DOCUMENT RESUME

ED 392 149 EA 027 378

AUTHOR	Gaustad, Joan
TITLE	Assessment and Evaluation in the Multiage Classroom. Special Issue.
INSTITUTION	Oregon School Study Council, Eugene.
REPORT NO	ISSN-0095-6694
PUB DATE	96
NOTE	82p.
AVAILABLE FROM	Oregon School Study Council, 1787 Agate Street, College of Education 5207 University of Oregon, Eugene, OR 97403-5207 ($10, nonmembers; $7.50, members; quantity discounts).
PUB TYPE	Collected Works - Serials (022) -- Information Analyses (070)
JOURNAL CIT	OSSC Bulletin; v39 n3-4 spec iss Jan-Feb 1996
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Academic Achievement; Classroom Techniques; Competence; *Educational Assessment; Elementary Secondary Education; *Mixed Age Grouping; *Nongraded Instructional Grouping; Performance; Portfolio Assessment; *Student Evaluation; *Testing Problems
IDENTIFIERS	*Authentic Assessment

ABSTRACT

Assessment of student progress is a challenge for educators who use developmentally appropriate practices such as multiage grouping. Interest in alternative types of assessment has become widespread. These performance-based, or authentic, assessments are explored in this document, which presents assessment strategies that work effectively with multiage instructional approaches. Chapter 1 begins by examining the purposes of assessment and then compares the characteristics, strengths, and limitations of conventional and authentic assessments. Chapter 2 explores methods used to assess and document the process of learning, such as observation, anecdotal records, and developmental checklists, and presents means of assessing, evaluating, and organizing authentic products of student learning. Issues involved in reporting student progress to parents and administration are examined in the third chapter. Chapter 4 considers the implications of authentic-assessment approaches for administrators and school boards, and summarizes what administrators should know about teachers' requirements to effectively implement new assessment methods. A summary publication is included. The appendix contains an overview of authentic-assessment practices in Oregon. Data were gathered from interviews with 10 educators and assessment specialists. (Contains 97 references.) (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 392 149

ASSESSMENT AND EVALUATION IN THE MULTIAGE CLASSROOM

Joan Gaustad

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
P. Pseke
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EA 027 378

Special Issue

Oregon School Study Council
January and February 1996 • Volume 39, Numbers 3 and 4

OSSC BULLETIN

Assessment and Evaluation in the Multiage Classroom

Joan Gaustad

Special Issue

Oregon School Study Council
January and February 1996 • Volume 39, Numbers 3 and 4

ISSN 0095-6694
Nonmember price: $10.00
Member price: $7.50
Quantity Discounts:
10-24 copies - 15%
25-49 copies - 20%
50+ copies - 25%

OSSC STAFF

Philip K. Piele, Executive Secretary
Stuart C. Smith, Editor
Linda S. Lumsden, Associate Editor
C. Joelle Serbacic, Layout Specialist
Juan Ferret, Publications Sales

OSSC GOVERNING BOARD

Bart McElroy, School Board Member, Salem/Keizer SD 24J
Joan Peterson, School Board Member, Ontario SD 8
Ronald Russell, Superintendent, David Douglas SD 40
Elaine Hopson, Superintendent, Tillamook SD 9
Marilyn Pheasant, Superintendent, Olney SD 11
Philip K. Piele, Professor and Executive Secretary,
Oregon School Study Council
David Conley, Associate Professor, Educational Policy and Management,
University of Oregon
Cliff Kuhlman, School Board Member, Oregon School Boards Association

OREGON SCHOOL STUDY COUNCIL

1787 Agate Street
College of Education
5207 University of Oregon
Eugene, Oregon 97403-5207
(541) 346-5044
Fax: (541) 346-2334

The University of Oregon is an affirmative action, equal opportunity employer.

Preface

As educators across the continent implement developmentally appropriate practices such as multiage grouping, integrated instruction, and continuous progress, they find that forms of assessment designed to work with conventional age-graded practices become less effective and more difficult to use as instruction changes. Teachers need new ways to assess, evaluate, and report student progress when their students vary in age and move at different rates toward individualized goals instead of marching toward uniform grade-level standards as in the past.

Meanwhile, even within traditional age-graded systems, dissatisfaction with the limitations of conventional assessments—particularly standardized testing practices—has given rise to intense interest in alternative types of assessment. Various types of performance assessment are being vigorously discussed, researched, and implemented at local, state or province, and national levels.

These alternative, or *authentic*, assessment approaches are the subject of this Bulletin. While its special focus is the multiage context, particularly at the primary level, most of the approaches it describes can also be used in age-graded classrooms, as well as in transitional classrooms that are adopting elements of developmentally appropriate practices and stretching, if not yet abandoning, traditional graded organization.

Joan Gaustad, the author of this Bulletin, received her B.A. in Psychology from Grinnell College in Grinnell, Iowa, and her M.A. in Clinical Psychology from John F. Kennedy University in Orinda, California. She currently works as a freelance writer in Eugene, Oregon. She explored non-age-graded instruction and the pitfalls encountered in implementing it in several previous OSSC Bulletins: *Nongraded Education: Mixed-age, Integrated, and Developmentally Appropriate Education for Primary Children* (March 1992); *Making the Transition to Nongraded Primary Education* (April 1992); and *Nongraded Education: Overcoming Obstacles to Implementing the Multiage Classroom* (Special Issue, November and December 1994).

Contents

	Preface	iii
	Introduction	1
	A Few Definitions	2
	Authentic Assessment and Nongraded Education	4
	Trends in Assessment	5
	An Overview of This Bulletin	6
1.	**What Is Good Assessment?**	7
	The Purposes of Assessment	7
	Conventional Assessments	9
	Authentic Assessment	17
	Conclusion: Balancing and Combining Assessments	22
2.	**Assessing Student Learning: Products and Processes**	24
	Setting Clear, Assessable Objectives	24
	The Tools of Authentic Assessment	27
	Involving Students in Assessment and Evaluation	33
	Conclusion	37
3.	**Reporting Student Progress to Parents**	38
	What! No Grades?	38
	Ongoing Communication with Parents Is Crucial	39
	Progress Compared to What?	41
	Written Progress Reports	44
	Conferences with Parents and Children	47
	Conclusion	50
4.	**Implications for School Boards and Administrators**	52
	Helping Teachers Implement Authentic Assessment	53
	Educating and Involving the Community	56

Conclusion 58

Appendix: Authentic Assessment in Oregon 59

Bibliography 62

Interviews 69

Other Multiage Titles Available from OSSC 70

Introduction

Primary teacher Terry Snyder had four years of multiage teaching experience in various subject areas under his belt when he decided to plunge into mixed-ability mathematics. An amateur jazz musician relatively comfortable with improvisation, he was the first member of the five-person Westmoreland Elementary School Primary Team* to abandon the traditional ability-homogeneous math groups.

Snyder used his classroom budget to purchase manipulatives, scales, and other measuring devices instead of texts and workbooks. He engaged his six- through eight-year-old students in an ever-changing variety of concrete, hands-on math tasks. Partnered with different classmates each day, students at different functional levels often practiced math skills of varying complexity while engaged in the same activities. A child still mastering number order might count and stack cubes "1-2-3-4," while his partner practiced her times tables, counting by groups of two, three, or ten cubes.

Students practiced addition by rolling dice on the carpeted floor and writing down the numbers for their partners to solve. Occasionally Snyder duplicated workbook pages for drill and practice. "The kids loved it," he related. "They *cheered* the first time I told them I was going to give them a times test, and they could do it with paper and pencil, and *they* didn't have to write the problems!"

Many activities integrated math and science. "We charted all kinds of things. We made birthday charts, from oldest to youngest. We made height charts. We made blue-eyed, green-eyed, red-eyed charts." The "ball roll" was especially popular: rolling different balls down an inclined plane in the breezeway, measuring the distance to the spot each one stopped, and charting the results.

Snyder was pleased with the success of his approach. Children appeared to be learning skills and concepts at least as well as in the past, while

*See the April and May 1992 and November/December 1994 OSSC Bulletins for a more detailed description of the Westmoreland primary program.

developing interpersonal skills and a wonderfully positive attitude toward math. Everything seemed to be going splendidly—until the end of the first reporting period approached. Then, Snyder recalled, "It was Panicsville for a few days."

"I had all kinds of observational data in my mind, but I didn't have the formal documentation because I had gotten away from a canned system where it was built in," he explained. "I'd seen them demonstrating competency, but I didn't have my traditional packets of papers saved, with the nice little percentages and marks and so on. And a lot of it didn't show on paper. It was things that you did with your body, and that you manipulated."

How Snyder ultimately solved his reporting problem is described in a later chapter. This Bulletin is intended to help educators successfully negotiate such pitfalls by presenting assessment strategies that work effectively with multiage instructional approaches.

A Few Definitions

Before exploring these new types of assessment, it may be helpful to review some basic assessment terminology.

Assessment and Evaluation

The terms *assessment* and *evaluation* are often used interchangeably. But although they can be considered part of a single process, they are technically distinct steps. The British Columbia Ministry of Education defines *assessment* as "the systematic process of gathering evidence of what a child can do," while *evaluation* is the process of interpreting that evidence and making judgments and decisions based upon that interpretation (Ministry of Education 1990b). Readers should be aware of this distinction. However, for the sake of brevity this Bulletin will sometimes use the term *assessment* to refer to the entire process.

Reliability and *validity* are key concepts in evaluating assessment quality. *Reliability* means the degree to which scores or ratings are consistent or dependable, "the degree to which test scores can be attributed to actual differences in test takers' performance rather than to errors of measurement" (National Association for the Education of Young Children 1988). For example, results obtained should ideally be the same when a student is assessed with the same test on different occasions, with different tests designed to measure the same thing, or when an assessment is rated or scored by different individuals. *Validity* means the extent to which an assessment actually measures what it is intended to measure.

Different types of assessment are appropriate at different times. Class-

room teachers use *formative* assessment on an ongoing basis to plan and modify instruction. *Summative* assessment is carried out at the end of a time period, to determine overall student accomplishment to date or to ascertain the effectiveness of a program. *Diagnostic* assessment, an indepth evaluation process used to identify children with special needs, is conducted when the existence of a serious problem is suspected, often by a team of professionals (NAEYC 1988; NAEYC and National Association of Early Childhood Specialists in State Departments of Education 1991; George F. Madaus and Thomas Kellaghan 1993; National Center for Research on Evaluation, Standards, and Student Testing [NCRESST] undated a). This Bulletin focuses on formative and summative classroom-level assessments conducted by teachers.

Nonconventional Assessment Approaches

"To many people, the words 'assessment' and 'testing' evoke the same image: rows of desks with students sitting silently working on paper-and-pencil tasks, perhaps filling in bubbles or circling responses to short questions concerning isolated snippets of information," writes Marianne Lucas Lescher (1995). *Authentic assessment*, *alternative assessment*, and *performance assessment* are basically equivalent terms for an approach that developed in reaction to such practices. A common factor in such assessments is that students typically

> generate rather than choose a response. Performance assessment by any name requires students to actively accomplish complex and significant tasks, while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems. Exhibitions, investigations, demonstrations, written or oral responses, journals, and portfolios are examples. (Joan L. Herman and others 1992)

Alternative Assessment. According to Vito Perrone (1991), *alternative assessment* is the label most educators use for these methods. Like *nongraded* (see below), this term defines itself negatively, with reference to what it is *not* rather than what it *is*. Perrone comments that he dislikes the term because it "gives too much legitimacy to the processes currently dominating assessment in schools."

Performance or *Performance-based Assessment.* The adjective *performance* emphasizes that these methods require student actions other than written responses to abstract questions. It is frequently used in the literature concerning assessment changes being proposed in traditional age-graded contexts, including large-scale standardized assessments being designed to replace standardized multiple-choice tests.

Authentic Assessment. The adjective *authentic* emphasizes that these methods are applied within the normal classroom context, or in settings that attempt to reproduce typical learning experiences, rather than in artificial contexts unlike those in which skills and knowledge are used and applied. "Authentic evidence is evidence that predominately: is selected in terms of program goals and learning experiences, reflects the regular conditions of the classroom, documents growth in children's actual 'products' rather than on work substitutes in contrived tasks; reflects some kind of real-life purpose, meaning or validity" (Ministry of Education 1990b). *Authentic assessment* is the term most commonly used in the literature focusing on nongraded or multiage instructional contexts, and the one I will generally use in this Bulletin.

Authentic Assessment and Nongraded Education

In previous Bulletins, I applied the term *nongraded education* to a group of innovative, overlapping educational practices that share a common research base and many elements of a common philosophy. These practices, whose implementation has become increasingly widespread in recent years, include non-age-graded organization, mixed-age and multiage grouping, integrated instruction, and authentic assessment.

Mixed-age and *multiage grouping* are terms for the practice of teaching children of different ages in the same classroom. Strictly defined, mixed-age grouping includes children ranging in age by up to two years, while multiage grouping includes children ranging in age by more than two years. However, the terms mixed-age and multiaged are often used indiscriminately. According to Anita McClanahan, coordinator of Early Childhood Education for the Oregon Department of Education, well over four hundred Oregon schools are currently implementing mixed-age or multiage programs.

Developmentally appropriate practices identified by the National Association for the Education of Young Children and other early childhood researchers include all the practices listed above. The concept of *developmental appropriateness*, as articulated by the NAEYC , has two dimensions: (1) age appropriateness, or appropriateness with reference to the typical development of children as established by research; and (2) individual appropriateness, or responsiveness to individual differences in rate and pattern of growth, learning style, personality, and family background (NAEYC 1987). *Developmentally appropriate practices* is an accurate and inclusive term, but its length can make it awkward to use.

In my March 1992 Bulletin, attempting to clearly distinguish between non-age-graded organization and nontraditional forms of assessment, I wrote:

> *Nongraded education* is the practice of teaching children of different ages and ability levels together in the same classroom, without dividing them or the curriculum into steps labeled by "grade" designations . . . . [T]hose unfamiliar with [the term] often assume it means *not giving letter grades* rather than *not sorting children by grade levels*. While the use of alternative types of evaluation is usually part of the nongraded approach, it is only a small element. (Joan Gaustad, March 1992)

In fact, however, authentic assessment is a vital and integral part of the nongraded approach, whose basic tenet is that individuals learn differently and should not be subjected to identical treatment in the classroom. It is impossible to adapt instruction to individual needs without determining each student's strengths, weaknesses, interests, and learning styles, and monitoring both academic progress and progress in areas typically ignored by conventional assessments.

Nongraded is also a problematic term because it has been used with significantly different connotations by different authors and educators. In this Bulletin, which frequently refers to "grading" in the sense of "symbolic marking systems" (Robert H. Anderson and Barbara Nelson Pavan 1993), I will try to reduce confusion by avoiding the term altogether. Instead, I will use *developmentally appropriate practices* or the shorter *multiage practices* to refer to the range of nontraditional instructional approaches typically used in mixed-age and multiage classes.

Trends in Assessment

In recent years, alternative forms of assessment have been gaining in popularity in the United States and in Canada. The number of states that reported using performance assessments rose from seventeen in the 1991-92 school year to twenty-five in the 1993-94 school year (Karen Diegmueller 1995b). A recent survey of state writing-assessment programs determined that portfolio assessments were under consideration in almost half of the states (National Center for Education Statistics, January 1995b). National organizations such as the National Writing Project are researching and disseminating information about alternative assessments, and the National Association for Educational Progress uses performance-assessments to assess student writing and reading proficiency across the nation (NCES, January 1995a and b).

This advancing tide has not gone unchallenged. In several states assessment changes have encountered opposition. For example, the pioneering California Learning Assessment System was suddenly terminated in fall 1994 after the election of a group of legislators opposed to its implementa-

tion (Diegmueller 1995b). Nonetheless, the general trend seems clear. Soon it may be hard to find students who have *not* encountered some form of alternative assessment during their school careers.

Merely because an approach is popular does not mean policy-makers should adopt it. However, policy-makers would be wise to encourage local educators to explore authentic assessment. In some cases, such as in Kentucky, state-mandated programs have set unrealistic deadlines for implementing authentic assessment, and schools that had previously relied strictly on conventional methods found themselves struggling to catch up. Laying some groundwork in advance may be advantageous, whether or not such assessments are ever required.

An Overview of This Bulletin

Chapter 1 begins by examining the purposes of assessment, then compares the characteristics, strengths, and limitations of conventional and authentic assessments.

Chapter 2 explores methods used to assess and document the process of learning, such as observation, anecdotal records, and developmental checklists or continua, and presents means of assessing, evaluating, and organizing authentic products of student learning.

Chapter 3 examines issues involved in reporting student progress to parents and administration. Chapter 4 concludes by considering the implications of authentic-assessment approaches for administrators and school boards, and summarizes what administrators should know about teachers' requirements to effectively implement new assessment methods.

Chapter 1

What Is Good Assessment?

The question "What is good assessment?" has many answers. Each type of assessment has strengths and weaknesses, and no single assessment method can serve all needs equally well. Which characteristics are most important depends on the purpose the assessment is to serve and the audience that will ultimately use the information it produces. To select appropriate types of assessments, educators must be clear about these purposes and understand how different methods match those purposes.

The Purposes of Assessment

Two main purposes of assessment and evaluation are to support student growth and learning and to facilitate accountability. While these purposes are not mutually exclusive, their requirements sometimes conflict.

Different Purposes Serve Different Audiences

Connie A. Bridge and her colleagues (1992) suggest considering these purposes in terms of the different audiences that are served. The primary purposes of assessment for these various audiences form a continuum, "with assessment for learning at one end and assessment for accountability at the other." Some of these audiences and purposes are presented below, beginning at the "learning" end of the continuum.

For students, principal purposes of assessment include providing corrective and reinforcing feedback on their progress, helping them develop self-evaluative skills, and stimulating pride in their achievements. Assessment also communicates to students what types of skills and knowledge their schools and communities value.

For teachers, assessment's many purposes include identifying students' strengths, weaknesses, and learning styles, and pinpointing skills and

concepts already mastered and those requiring more time and practice. This information helps teachers plan curriculum and instruction, set appropriate individual goals for students, and provide extra help or additional challenges. Assessment also provides feedback on the comparative effectiveness of instructional strategies.

For parents, the most important purpose of assessment and evaluation is to provide information on their children's progress. Assessment can alert parents to special talents that should be fostered as well as areas where extra support is needed, and it may suggest ways they can support their children's learning.

For administrators and policy-makers, assessment facilitates program evaluation, determination of staff development needs, and planning or modification of school-improvement programs.

For legislators and other public officials, assessment serves accountability purposes. Comparisons of student achievement and program effectiveness permit identification of successful programs that deserve acknowledgment and emulation, help target schools and districts where improvement is needed, and guide the distribution of resources to support progress.

High-Stakes and Low-Stakes Assessments

Assessments can also be considered in terms of the significance of the consequences that may result from their use—how high the "stakes" are.

Assessments used as the basis for decisions that have major effects on students, educators, schools, and other stakeholders are often called *high-stakes* assessments. High-stakes assessments usually occur infrequently, and results are interpreted without background knowledge of the child or classroom circumstances. This means that each assessment carries great weight, that errors or misleading results are difficult to identify, and that much time is likely to pass before decisions made and actions taken on the basis of erroneous information can be reversed. The quality and accuracy of high-stakes assessments are therefore extremely important.

By comparison, classroom assessments used for learning purposes are relatively *low stakes*. Because teachers use a variety of assessments on an ongoing basis, no single assessment has overwhelming weight. Teachers can interpret results in the context of classroom life and their personal knowledge of the student. If erroneous judgments are made and inappropriate actions taken due to atypical or unrepresentative results, errors are likely to be quickly discovered and actions taken to remedy problems. As Lorrie A. Shepard (1989) points out, although classroom assessments "are probably less reliable (in a statistical sense) than a one-hour standardized test, the accumulation of data gathered about individual pupils has much more accuracy."

Conventional Assessments

- Paper-and-pencil true-false, fill-in-the-blank, and multiple-choice tests.
- Percentage and letter-grade evaluation: 90 percent or above is an A, 80 to 89 percent is a B.
- Grading on a "curve."
- Computer-scored standardized tests taken by multitudes of students across the nation.

These are the familiar, traditional forms of assessment most of today's adults recall from their elementary-school days. They may be comfortingly familiar to adults, but the longer they have been used, the more evidence has accumulated concerning their negative effects on student learning.

Most people would probably be surprised to learn that these "traditional" assessments have a relatively short history. This section briefly surveys the history of conventional assessments and summarizes their major shortcomings, which have been described in depth by many other authors.

The History of Conventional Assessments

Oral recitation was the usual means of assessment during the thousands of years when literacy and formal education were reserved for an elite few. "When the class is reasonably small and the required knowledge sufficiently spelled out, student learning can be reliably monitored on the basis of daily recitation; it is clear which students have learned their times tables, have memorized twenty lines of poetry, and can list the kings, presidents, or dynasties in order" (Howard Gardner 1991). Written essay examinations joined the assessment repertoire in medieval times, and along with oral examinations, continued to dominate educational assessment throughout the world until the early twentieth century (W. James Popham 1993).

Schooling in colonial America largely conformed to the description above. But by the mid-1800s, the Industrial Revolution was transforming both society and education. "The factory system superseded the craftsman, bringing to industry the mass production of the assembly line. Meanwhile, growing confidence in the capacity of the human race for unending progress, the diffusion of religious humanitarianism, the beginning of the labor movement, and growing nationalism created a milieu that was receptive to the revolutionary idea of education for all" (John I. Goodlad and Robert H. Anderson 1987).

A new organizational system was needed to handle these unprecedented numbers of students. Age-graded organization was introduced to the

United States from Prussia in 1843, and it quickly became the standard throughout the nation. The new, efficient graded schools were proudly compared to factories.

Changes in reporting and assessment lagged behind at first. It wasn't until after the turn of the century that some high school teachers began using percentages to report student achievements. At this point elementary teachers were still documenting student learning with written descriptions. By 1918, some teachers had begun using a three-point rating scale consisting of Excellent, Average, and Poor, while others were using the five-point ABCDF rating scale destined to become standard. But while these scales superficially appeared more standardized and quantitative than written descriptions, they still relied on highly variable, subjective teacher judgments. As early as 1912, studies appeared challenging the reliability of percentage grading (Thomas R. Guskey 1994a).

Standardized tests, prepared by experts and tested for statistical validity and reliability, promised greater consistency and objectivity than teacher judgments, while their efficiency and economy matched that of graded organization. The first such test to be widely used in United States public schools, the *Thorndike Handwriting Scale*, appeared in 1909, and it was soon followed by others (Vito Perrone, Spring 1991). The U.S. Army's large-scale use of multiple-choice tests to assess and place recruits during World War I had a major impact on educational assessment. Multiple-choice tests spread rapidly after the war (Popham; Walter Haney and George Madaus 1989), and by the 1930s most schools in the U.S. and Canada used some type of standardized testing (Perrone, Spring 1991).

At first standardized testing was only an infrequent adjunct to classroom grading. Educators continued to seek ways to reduce the subjective nature—or at least appearance—of the more commonly used types of evaluation and reporting. Starting in the 1930s, when intelligence-test research suggested that intelligence in the general population was distributed along a bell-shaped probability curve, classroom grading "on the curve" became increasingly popular (Guskey 1994a). But many educators and citizens continued to distrust the accuracy and objectivity of teacher-assigned grades as compared to standardized test scores. "Test scores have the aura of scientific respectability and rigor, whereas teachers' judgments seem subjective and open to multiple sources of bias" (Scott G. Paris and Others 1991).

This faith in the superior accuracy of standardized testing contributed to its increasingly frequent use as the decades passed. Students completing high school in 1991 had taken eighteen to twenty-one standardized tests during their school careers (Perrone, Spring 1991).

The Limitations of Conventional Classroom Assessments

Inconsistency in grading among teachers has remained a major complaint concerning conventional classroom assessment and evaluation. For example, in 1989 Stiggins, Frisbie, and Griswold found that "different teachers in the same building sometimes adopted different cutoff scores for the same grade, or even used different reporting schemes for the same course" (Marcia M. Seeley 1994). Teachers also make different choices concerning how much weight, if any, should be given to nonacademic factors such as effort. Such inconsistency is practically guaranteed by the isolation of teachers in self-contained classrooms that is characteristic of age-graded organization. Teachers whose interaction with colleagues is limited to a few minutes of conversation over a cup of coffee are unlikely to develop consensus about assessment and evaluation criteria.

The subjective nature of assessments used by teachers can be an advantage, as Thomas R. Guskey (1994b) points out: "Because teachers know their students, understand various dimensions of students' work, and have clear notions of the progress made, their subjective perceptions may yield very accurate descriptions of what their students have learned." However, teacher bias can also distort perceptions of student performance. Factors such as disciplinary infractions, neatness, and cleanliness have been found to significantly affect teacher judgments concerning achievement, especially for boys (Guskey 1994b).

These problems might be reduced if most teachers received high-quality training in assessment and reporting, suggests Guskey (1994b). But few accredited teacher-education programs require assessment courses for graduation, even fewer states require assessment training for teacher certification, and the optional courses available have traditionally focused on large-scale, paper-and-pencil formats of little use to classroom teachers. Administrators generally cannot help the teachers they supervise or judge the effectiveness of their assessments because they are even less assessment-literate. As a result, many teachers have relied on prepared tests from textbooks or teachers' manuals rather than constructing their own (Richard J. Stiggins 1991).

Conventional reporting methods based on comparing student progress to that of classmates have inherent problems much more serious than inconsistency. Robert E. Slavin (1986) paints a poignant picture of the psychological torture experienced by slower students in a competitive classroom environment. ABCDF reporting may be considered desirable by students who are "winners" in the grading game—and competitive parents of such students who enjoy exulting in their children's superiority. But such grading has been found to negatively affect the motivation and quality of work of able students

as well as less able ones, "with the most destructive effects occurring in activities that require creativity or higher-order thinking" (Alfie Kohn 1994).

Grading on a curve is even more problematic. "The bell-shaped curve is used for statistical convenience, not because any form of knowledge or ability is actually distributed in this manner," contend D. Monty Neill and Noe J. Medina (1989). Even if ability were so distributed, the number of children in one classroom is far too small a sample for use of a curve to be statistically valid (Ministry of Education 1995b). Artificially limiting the number of good grades and requiring some students to get Ds and Fs increases competition among students, destroys collaboration and community, and damages the self-esteem and motivation of students unfortunate enough to learn more slowly than their classmates (Guskey 1994b, Kohn 1994). On the other hand, scoring high on the curve says nothing about absolute quality of achievement, only relative quality. A student can earn high grades with mediocre work if most of his or her classmates do even more poorly.

The Strengths and Weaknesses of Standardized Tests

The NAEYC defines standardized tests as assessment instruments "that are composed of empirically selected items; have definite instructions for use, data on reliability, and validity; and are norm- or criterion-referenced." *Norm-referenced* test scores are reported in terms of how the test-taker's performance compares with that of other test-takers, whereas *criterion-referenced* tests report performance in relation to specified performance levels. Intelligence tests, achievement tests, developmental screening tests, and diagnostic assessment tests are all types of standardized tests (NAEYC 1988).

Standardized tests are unsurpassed in terms of efficiency and economy (Blaine R. Worthen 1993a). Due to the huge numbers of students involved, test creators can invest sufficient resources to ensure statistical validity and reliability far surpassing that of teacher-created tests. Administration of the tests takes comparatively little time, and thousands of multiple-choice response sheets can be rapidly scored by high-speed electronic scanning machines (Popham). No other form of assessment can equal well-designed standardized tests for overall comparisons of the performance of large groups of students (Blaine R. Worthen and Vicki Spandel 1991).

The comparative information provided by nationally used norm-referenced tests is more accurate and useful than comparisons involving the few students in one school or classroom. However, such information still has only limited value, as Goodlad and Anderson point out:

> Helpful as it is to know that a given proportion of all children of approximately the same age are better or poorer at a given task than

> the child we have in mind, it is far more valuable to know how his performance compares with his own past performance, what appear to be the direction and rate of his development in mastering tasks in that field, and how well this performance relates to what the teacher has planned for him to do.

Standardized test results are only reliable and valid, and their results comparable, when the tests are administered under standard conditions, to appropriate populations, for the purposes for which the test was designed (NAEYC 1988). Altering any of these variables can invalidate test results. The best tests inevitably contain margins of error, even when conducted under ideal conditions. And while some tests may accurately predict the future performance of *groups* of students, predicting *individual* performance is quite another matter (Worthen and Spandel).

Unfortunately, test users have seldom understood these limitations. As was previously mentioned, most teachers and administrators lack adequate professional preparation in assessment (Stiggins 1991), and parents, school board members, public officials, and the general public are even less knowledgeable. Test users have often maintained blind faith in so-called "objective" tests, ignoring the fact that test questions are created and selected by fallible human beings.

All too often, cautions against inappropriate use have remained "buried in hard-to-read manuals" (Neill and Medina) while questionable test results were used as the basis for significant decisions, many of which have long-lasting negative emotional effects on those tested. Students have been selected or barred from programs because they missed arbitrary cutoff scores by a few points on tests that have much larger standard deviations. Children have been deemed not ready to start first grade because of tests with predictive validities so low that up to 50 percent of children could be misidentified (Neill and Medina). Small, statistically insignificant variations in test scores between districts, or within the same district from year to year, have caused jubilation or brought the wrath of the public down on the heads of hapless teachers and administrators.

Many educators responded to these pressures by focusing instruction on the specific skills and formats that occurred in the tests, to the detriment of other skills, content areas, and formats. For example, some teachers stopped using essay tests "because they were inefficient in preparing students for multiple-choice tests" (Shepard 1989). In some districts efforts to increase test scores included unethical practices ranging from drilling students in advance on actual test items to sending low-achieving students on field trips during the week of testing. Thomas M. Haladyna and others (1991) describe the extent of test-score polluting practices as "staggering." Widespread use of such practices resulted in further reductions in test reliability and validity.

In the short run, such tactics sometimes resulted in higher test scores, but they had negative long-term effects on student learning, especially higher order learning, as well as causing anxiety, decreased motivation, and damaged self-esteem.

> Since about 1970, when standardized tests began to be used for a wider variety of accountability purposes, basic skills test scores have been increasing slightly, while assessments of higher order thinking skills have declined in virtually all subject areas. Officials of the National Assessment of Educational Progress, the National Research Council, and the National Councils of Teachers of English and Mathematics, among others, have all attributed this decline in higher order thinking and performance to schools' emphasis on tests of basic skills. (Linda Darling-Hammond, 1993)

Complaints about standardized testing and calls for testing reform grew louder as the results of such misuse became evident. The National Association for the Education of Young Children and the Association for Childhood Education International (ACEI) published position statements criticizing the use of standardized tests with young children, who are particularly vulnerable to the damaging practices fostered by testing pressure as well as being erratic test-takers (NAEYC 1988; Perrone, Spring 1991). Some criticism focused on fairness issues. Critics charged that test questions assume background knowledge and vocabulary associated with mainstream, white, upper-middle-class experience and are biased against students of different cultural, linguistic, and socioeconomic backgrounds—students further disadvantaged by inferior educational experiences and less test-taking preparation. Some critics go further, charging that neither standardized tests nor other conventional assessments measure the most significant aspects of learning for *any* student.

Do Conventional Assessments Measure What Is Important?

Outside the classroom, the ability to produce isolated skills and recall bits of abstract knowledge on demand is much less important than being able to apply skills and knowledge appropriately. Adults solving real-world problems

> have to analyze, synthesize, interpret, and evaluate facts and ideas far more often than they have to "know" them in the sense of only being able to recite them. These are precisely the kinds of thoughtful abilities that more and more leaders say graduates should possess and that fewer and fewer graduates actually do possess. And they are precisely the activities that cannot be tested the way we currently test students in most schools. (Rexford Brown 1989)

Howard Gardner cites dismaying evidence that such higher order cognitive abilities are neither being learned by students nor measured by conventional assessments. An "overwhelming body of educational research" demonstrates that students with good grades and high standardized test scores typically don't understand what they have studied. College honor students with physics and engineering training exhibit primitive misconceptions comparable to those of ten-year olds when confronted with physics problems outside the standard "correct-answer" test format. Students harbor elementary misunderstandings concerning concepts as basic as evolution and the laws of heredity even after two years of biology coursework. College students who can produce sophisticated explanations of past events for history and economics exams revert to crude simplifications and stereotypes when asked to explain current events.

In mathematics, students of all ages display what Gardner calls "the practice of *rigid application of algorithms* . . . .[W]hen given a string of numbers, students immediately and reflexively begin to perform certain operations upon them." Correct answers result as long as word order in the problem parallels the order of symbols in the equation, but if the problem is rephrased, the automatic machinery screeches to a halt.

Conventional assessments also ignore important nonacademic areas of competence such as the abilities to work cooperatively with others, to make independent decisions, and to effectively self-evaluate.

Conventional Assessments Are Based on Outdated Assumptions

Conventional assessment methods assume that learning involves memorizing discrete chunks of objective information in a linear, step-by-step fashion; that learning can be effectively assessed by demanding that desired chunks be produced in timed, paper-and-pencil tests; that academic achievement can be measured and quantified by counting correct responses; and that nonacademic areas where progress cannot be quantified are not worth attempting to assess. These assumptions were derived, at least in part, from the behaviorist theories of learning that dominated psychology in the early part of the century.

Decades of research in cognitive psychology and child development have disproved these assumptions. While learning is far from being completely understood, we now know that it is a complex, multidimensional process subject to great individual variation. Children "construct" knowledge, modifying previous understandings as they interact actively with their environment and other people, rather than absorbing knowledge passively. Individuals vary greatly in rate and pattern of cognitive development and rely on different learning styles, and the learner's emotional state affects learning

to a far great degree than was previously suspected.

Research by Howard Gardner suggests that intelligence, rather than being an entity measurable on a single scale, consists of at least seven *multiple intelligences* that are present to different degrees in different learners. Spatial, musical, kinesthetic, interpersonal, and intrapersonal intelligences exist in addition to the verbal-linguistic and logical-mathematical intelligences toward which education traditionally has been oriented.

Not only are conventional assessment practices restricted to measuring "rote, ritualistic, or conventional performances" (Gardner), they actively interfere with higher order learning and creativity. Renate Nummela Caine and Geoffrey Caine (1994) have found that the vast majority of students "downshift" to a limited, inflexible, lower order level of brain functioning when the following conditions exist in the classroom: prespecified "correct" outcomes are established by an external agent in the classroom, personal meaning is limited, rewards and punishments are externally controlled and relatively immediate, restrictive time lines are given, and work to be done is relatively unfamiliar with little support available.

As awareness spread concerning this new information on learning, educators across the continent began adopting research-based, developmentally appropriate instructional practices such as multiage grouping, integrated instruction, and cooperative learning with the aim of creating classroom environments that facilitate creativity and higher order brain functioning: environments in which outcomes are open-ended, learning is personally meaningful, intrinsic motivation is emphasized, and tasks are manageable and supported with relatively open-ended time limits (Caine and Caine).

Educators of young children led the way in promoting such practices. The NAEYC's *Developmentally Appropriate Practice in Early Childhood Programs Serving Children from Birth Through Age 8* (1987) was influential in disseminating updated knowledge about child development and learning, identifying developmentally inappropriate practices (including the use of letter grades during the primary years), and encouraging the implementation of appropriate practices.

But assessment practices were often slower to change than instructional and organizational practices. Standardized tests remained a nearly universal requirement, and innovating teachers who saw their students' learning improve often found themselves in hot water with administrators if test scores did not show a concomitant, immediate increase (Mary Lee Smith 1991). Even teachers with supportive administrators found standardized testing stressful. "It was very painful to subject children to the tests," said one teacher at New York City's Bronx New School. "Some of my kids had grown tremendously through the course of the year . . . . The records I kept of them could show this progress. But I knew it wouldn't show up on the tests."

Testing was also demoralizing for students, said another teacher at the same school: "Sometimes I felt like all the growth in self-esteem and self-confidence that took place in the course of an entire year went down the drain in the two or three hours of taking the test" (Beverly Falk, September 1994).

Educators in some alternative programs rebelled against damaging assessment practices to the extent of refusing to systematically assess student progress in any fashion. The absence of assessment did not harm students who were learning well, says Gardner. However, "any educational institution must face the possiblility that it is *not* effective and must demonstrate a willingness to reflect, evaluate, and change course as often as proves necessary." The inability of alternative programs to present evidence of effectiveness gave many members of the general public the impression that students in such programs "were just having a good time and not mastering anything" (Gardner).

Wiggins (1993) sums up the assessment dilemma as follows:

> If I had to choose between, on the one hand, mickey mouse "gotcha!" tests with norm-referenced scoring and grading and, on the other hand, an absence of uniform tests and grades, I *might* go with the alternative schools; but it is a bad choice . . . we must think more carefully about how to balance the nurturing of diverse intellectual urges with the need for maintaining and inculcating standards—a quest for humane yet effective rigor, standards without mere standardization.

Authentic yet effective assessment methods were clearly essential if developmentally appropriate practices were to be successfully implemented and generally accepted. In 1991 the NAEYC joined the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) to issue "Guidelines for Appropriate Curriculum Content and Assessment in Programs Serving Children Ages 3 Through 8," a position statement that explained developmentally appropriate assessment practices in greater detail than the NAEYC's complementary 1987 publication. Among other things, the statement called for the prohibition of "group-administered, standardized, multiple-choice tests . . . before third grade, preferably fourth" (NAEYC and NAECS/SDE 1991).

Authentic Assessment

The label "authentic assessment" may be new, but some of the methods it comprises are very old. Good teachers have used "authentic" assessment techniques to monitor their students' learning since long before the invention of conventional methods. According to Joan L. Herman and her

colleagues (1992), "What is new about these assessments is that they make explicit and formal what was previously implicit and informal."

The Goals of Authentic Assessment

Researchers and educators involved in developing authentic assessment put a high priority on avoiding the weaknesses of conventional assessments, particularly of standardized tests. Rather than asking "What methods are efficient, economical, and easy to use?" and allowing the answers to this question to determine the design of assessments, they began by asking "What should assessment be and do?" They set ambitious goals—far surpassing the goals of conventional assessment—and only then began attempting to create assessment methods capable of achieving these goals. These major goal areas, six in number, can be summarized as follows:

1. *To assess truly important aspects of learning.* Authentic assessments aim to assess higher order thinking skills and problem-solving skills, not just memorization of facts and lower level, decontextualized skills; to identify genuine understanding and creativity, not just "easily counted (but relatively unimportant) errors" (Grant Wiggins 1989); and to assess processes as well as products. Complex, higher order processes and products may be difficult to assess objectively and harder to quantify, but authentic assessment proponents are willing to forgo exactness rather than to ignore important learning areas. "It is better to make a tentative, subjective decision about an important goal or stage of development . . . than an absolute, objective judgment about a trivial one," asserts the British Columbia Ministry of Education (1990b).

2. *To assess learning comprehensively.* Authentic assessments monitor growth and learning in social and emotional areas as well as in academic areas. For example, the British Columbia Primary Program (Ministry of Education 1990b) identifies three "learning dimensions" to be assessed: attitudes and dispositions, including self-confidence, curiosity, and cooperation; skills and processes, including thinking, communicating, problem-solving, and quantitative reasoning; and knowledge and understanding, which includes "factual, conceptual and procedural knowledge" of various content areas.

Grant Wiggins (1993) eloquently describes the importance of nonacademic growth in terms of the ultimate goals of education:

> It is our *attitude* toward knowledge that ultimately determines whether we become wise (as opposed to merely learned.)....We must make habits of mind—the intellectual virtues—central to our assessment.... It is not the student's errors that matter, but the student's responses to error; it is not mastery of a simplistic task that impresses, but the student's risk taking with the inherently complex; it is not thorough-

> ness in a novice's work that reveals understanding, but full awareness of the dilemmas, compromises, and uncertainties lurking under the arguments he or she is willing to stand on.

The comprehensive, multidimensional nature of authentic assessment provides greater opportunities for success to students with nonstandard learning styles, students whose best-developed "intelligences" are other than the traditionally valued verbal-linguistic and logical-mathematical intelligences.

3. *To be responsive to the needs of individual learners.* Herman and her colleagues remind us that the root of the word *assessment* means "to sit beside." Authentic assessment reaffirms the importance of "sitting beside" and interacting with the learner: determining what responses really mean, clarifying ambiguous questions, providing and receiving feedback as part of the assessment process. "The standardized test is disrespectful by design," says Wiggins (1989), because it "treats students as objects—as if their thought processes were similar and as if the reasons for their answers were irrelevant . . . equity requires us to insure that human judgment is not overrun or made obsolete by an efficient, mechanical scoring system." Assessment must be flexible in order to be fair to students of different cultural, linguistic, and socioeconomic backgrounds; to students with different learning styles; in fact, to any student with a unique, nonstandard perspective.

Flexible and responsive assessment depends on the informed judgment of the teacher. As Brian Cambourne comments, "the knowledgeable and experienced 'human as instrument' is more effective, accurate, credible and trustworthy than the traditional 'test-as-instrument,' especially when making sense of human behaviour" (Ministry of Education 1992).

4. *To positively affect instruction.* Aware that assessment powerfully influences what is taught and how, the creators of authentic assessments strive to make that influence positive. Assessment should be "congruent with and relevant to the goals, objectives, and content of the program" and should not "place children in artificial situations, impede the usual learning and development experiences of the classroom, or divert children from their natural learning processes," state the NAEYC and the NAECS/SDE (1991). This can be done by conducting assessment as normal classroom activities proceed, or by designing stand-alone assessment tasks that are themselves meaningful learning experiences.

5. *Validity and reliability.* Validity and reliability are still highly desirable goals, though they are not regarded as more important than the goals just listed.

6. *Practical feasibility.* Finally, none of these goals will be achieved if assessments require excessive teacher time and energy, or if teachers feel the information about students they yield is not worth the cost and time they require. Assessments must be practically feasible if they are to be used.

How Well Does Authentic Assessment Achieve These Goals?

It is too soon to know how well authentic assessment will ultimately succeed in meeting these ambitious goals. "Currently, most developers of the new alternatives (with the exception of writing assessments) are at the design and prototyping stages," Joan L. Herman (1992) observes; "few have yet collected data on the technical quality of their assessments or about their integrity as measures of significant student learning." To date, authentic-assessment approaches appear most successful in the goal areas where the shortcomings of standardized testing are the most obvious, and weakest in the areas where standardized testing is strongest: efficiency, economy, and statistical validity and reliability.

Considerable progress has been made in creating systematic means of judging complex, higher order tasks whose evaluation might at first seem "hopelessly subjective" (Gene I. Maeroff 1991). Consistent, accurate assessments can be achieved through the use of *rubrics*, sets of scoring guidelines that state the dimensions of performance being assessed, provide scales of values for rating those dimensions, and sometimes provide standards for judging performance (NCRESST undated b and c). Rubrics will be examined in chapter 2.

Writing assessment is the best-developed such area to date and can serve as a model for assessments of other types of complex performance. The National Association for Educational Progress (NAEP) reports that writing assessment has undergone dramatic change throughout the nation over the past decade, stimulated in part by the increasing popularity of process-writing approaches. Many states have implemented or are considering some form of performance-based writing assessment (NCES 1995b).

In Oregon, for example, statewide writing assessment relied on multiple-choice tests until 1978. Now, students select one of two possible topics and are given forty-five minutes per day for three consecutive days to write, revise, and polish a final draft (Oregon Department of Education undated). Raters assess their writing proficiency across six dimensions—ideas, organization, voice, word choice, sentence fluency, and conventions (spelling, grammar, punctuation, etc.)—using the Oregon Analytic Model, which was derived from a rubric developed by teachers from Beaverton and Portland school districts (Barbara Wolfe, Michael Dalton, and Wayne Neuberger 1993).

However, major problems of validity, reliability, and generalizability remain in assessing complex, higher order abilities. Inconsistency in performance among student learners is a particular problem for researchers designing large-scale performance assessments intended to replace large-scale multiple-choice tests. In one study of hands-on science assessment, researchers found that student achievement could be judged accurately only when ten

to twenty tasks were assessed (Worthen 1993a). "One-event testing in the performance area is even more dangerous than one-shot multiple-choice testing, because multiple-choice tests have many different but related items, which makes reliability easier to get and measure," explains Grant Wiggins (Ron Brandt 1992).

Collecting samples of student work over time in portfolios would seem an easier, less costly way to acquire multiple examples of student performance. However, Joan L. Herman and Lynn Winters (1994) note that little research has been done so far concerning the technical quality of portfolio assessment, and that existing data show uneven results. For example, consistency is hard to achieve in the key area of *interrater reliability*, or consistency in scoring of portfolios among different raters. "Available data suggest that such consensus depends on clearly articulated criteria, effective training, and rubrics that reflect shared experience, common values, and a deep understanding of student performance," they explain.

A major drawback of authentic-assessment methods is that they are more time-consuming and labor-intensive than conventional assessments. Teachers generally need considerable professional development to learn the theoretical foundations of authentic assessment and to develop consensus on assessment criteria, plus ongoing support as they gain practical experience learning to use the new assessment methods in the classroom. They also need extra planning time on a permanent basis to continue to use the methods effectively.

Providing this additional staff time is expensive, but essential for success. "Current calls for assessment-driven reform acknowledge the need for staff development but tend to underestimate the extent and depth of what is needed," comments Lorrie A. Shepard (1995). "While teaching toward open-ended tasks might be an immediate improvement over worksheets designed to mimic standardized tests, our experience shows that well-intentioned efforts to help kids improve at assessment tasks can be misdirected if teachers do not understand the philosophical and conceptual bases of the intended curricular goals."

It is worth emphasizing that many authentic-assessment techniques work poorly with traditional instructional methods, and thus their implementation must be part of a greater, comprehensive change. On the other hand, many teachers feel the additional information authentic assessments yield about students is well worth the extra time and effort.

Fortunately for multiage teachers, authentic assessment's most serious problems relate to large-scale assessments that seek to compare student achievement, while its strengths are most evident at the everyday, classroom level. Many educators report that in their personal experience, authentic assessment facilitates positive changes in instruction and student learning

(Herman and Winters). Teachers implementing multiage practices find authentic assessment methods a far better match for their teaching than conventional methods. "Schools should be quick to capitalize on alternative assessment, whenever appropriate, for it seems clear that it offers much at the local level," advises Worthen (1993a).

In a report released August 1995, a National Academy of Education panel on standard-based assessment reform suggested that high-stakes performance-based testing be postponed, but urged educators to begin using alternative assessments in the classroom immediately, despite their imperfections. "To wait until fully developed standards and assessment instruments are available in all knowledge domains, and capacity to implement them exists in all areas of the country, as opposed to acting on what we know now, would cheat many American students out of beneficial changes," said the panel's report (Karen Diegmueller 1995c).

The greatest dangers to authentic assessment may be the uncritical enthusiasm of some of its advocates and unrealistic expectations concerning the length of time it should take for this still-young assessment approach to achieve its potential. Worthen (1993a) warns that well-meaning proponents of alternative assessment who downplay its weaknesses could "raise stakeholders' expectations to unrealistic levels, thus leading to disappointment and ultimately the withdrawal of support." Given time, researchers and educators will undoubtedly refine existing authentic-assessment strategies and develop new, more effective ones. In the meantime, educators should maintain an open-minded but realistically critical attitude toward all types of assessment.

Conclusion: Balancing and Combining Assessments

In its current state of development, authentic assessment best serves the first purpose cited in this chapter: to support student growth and learning in the classroom. Carefully implemented authentic assessments can greatly benefit students, teachers, and parents in "low-stakes" situations. Teachers should try promising techniques with the awareness that a few "bugs" may need to be worked out, and that some techniques simply may not work. These low-stakes classroom assessments are explored in the remaining chapters of this Bulletin.

Authentic assessments still have serious drawbacks for high-stakes, accountability purposes. Research continues on ways to improve the validity and reliability of large-scale, standardized performance assessments, and on strategies to reduce costs while still producing accurate group achievement results, such as testing scientifically selected samples of students on different assessment tasks instead of testing every student for every assessment task (Worthen 1993a).

However, authentic assessments are not immune from the test-score corruption that can result when educators feel pressured to raise scores by any means possible. As long as test results are used for accountability purposes and affect things like financial support, school district certification, and teacher evaluation, assessment for learning will tend to be pushed into the background. "When the stakes are high, people are going to find ways to have test scores go up," concludes George Madaus (Ron Brandt 1989).

In the meantime, standardized tests remain valuable for some uses despite their limitations. Bredekamp comments in her review of the British Columbia Primary Program, "The virtues of standardization (systematic data collection, reliability, objectivity) do not need to be sacrificed and should not be neglected with the use of more authentic assessments such as teacher observation and portfolio review." Standardization in *assessment* can be very helpful, she maintains, particularly in identifying problem areas where children need more help; the error to avoid is allowing standardization to prevail in *evaluation*, the interpretation and decision-making parts of the process (Ministry of Education 1992).

Teachers, administrators, and policy-makers must keep in mind that no assessment method is perfectly accurate and that different assessments make different kinds of errors. "A critical attitude toward assessment and a wider appreciation of its effects on teaching and learning" may be more important than the merits of any particular assessment method, say Haney and Madaus. If they make themselves knowledgeable concerning the strengths and limitations of different assessment approaches, educators can select the most appropriate assessments for specific purposes and combine assessments with differing strengths to create a more balanced, multidimensional understanding of student learning than any single assessment tool could provide.

Chapter 2

Assessing Student Learning: Products and Processes

In the conventional classroom, assessment focusses primarily on the *products* of the learning process: student writings, drawings, test papers, completed worksheets, and so forth. In the multiage classroom, *processes* of student growth and learning are also assessed.

In fact, it is often difficult to separate product and process. Examining samples of a student's work over time reveals much about that student's learning process, and observing *how* the student goes about creating a piece reveals much about his or her skills and knowledge. Some student creations, such as demonstrations, skits, and oral presentations, contain elements of both. Many of the assessment methods described in this chapter can be applied to both products and processes.

Setting Clear, Assessable Objectives

Authentic assessment considers all aspects of children's growth and development. For example, the British Columbia Primary Program sets goals and assesses student progress in five areas: aesthetic and artistic development, emotional and social development, intellectual development, physical development, and development of social responsibility (Ministry of Education 1991). Before progress can be assessed in any of these areas, clearly stated, measurable objectives must be established, and the criteria for meeting those objectives must be agreed upon by educators and communicated to students and parents. Only then can instruction be planned and appropriate assessment tools selected to monitor progress toward those goals.

Establishing General Objectives

It is best to involve as many stakeholders as possible in defining major instructional goals. For example, when Fairplay Elementary School in Corvallis, Oregon, was in the process of changing their focus to improving student learning and defining what kids should know and be able to do. Fairplay staff invited parents, community leaders, and former Fairplay students, as well as teachers from Fairplay's feeder middle and high schools and professors from Oregon State University, to join them in developing new literacy and numeracy outcomes (Fairplay Elementary School, February 1994). "We wanted input from other folks so we weren't just recreating what already exists," explained Fairplay Principal Julie McCann.

A few brainstorming sessions can generate an overwhelming number of suggestions for good outcomes. "Don't feel that you will be able to cover all the standards you have discussed," NCRESST (undated b) advises educators reassuringly. "At this point you will need to prioritize your standards. Ask yourself what kids *must* know and what would be *useful* for them to know. Ask yourself what *you* can feasibly achieve with the resources currently available to you."

The British Columbia Ministry of Education has sought input from stakeholders at every step of its education reform process. The outcome statements that form the basis of the mandated provincial curriculum were written by teachers, in response to the advice of overview groups that included representatives of various education partners and specialist organizations. Outcomes, each of which begins *It is expected that students will...* were required to be "observable and reportable, as well as understandable by teachers, parents, students and the general public," and were subjected to a rigorous review process before adoption (Ministry of Education, September 1995).

Setting Specific Performance Criteria

A list of excellent outcomes is only the first step in a challenging process. "It is easier to propose outcomes than it is to set the criteria and establish the performance levels that are represented by various achievements," asserts Maeroff. Well-articulated criteria that "represent teachable and observable aspects of performance" (Herman and others) are an essential prerequisite to fair, consistent assessment. They clarify the goals of instruction, guide teacher planning, and communicate to students, parents, and the community what goals and values the school considers important. Herman and her colleagues present many excellent suggestions for creating clear, unambiguous, and unbiased performance criteria.

Developing criteria is a time-consuming, ongoing process, and educa-

tors should expect to do a lot of "fine-tuning" as criteria are put into practice. For example, scoring criteria developed at Mark Twain Elementary School in Littleton, Colorado, went through at least ten major revisions. "Every time we gave the assessment, we saw some student doing something we couldn't account for on the scoring rubric," said Principal Monte Moses (Maeroff).

British Columbia's *provincially prescribed learning outcomes* provide excellent models of clear, well-defined instructional objectives. As part of the province's ongoing education-reform process, the Ministry of Education is replacing the former array of curriculum documents in more than forty-seven different formats with *Integrated Resource Packages* in each content area (Ministry of Education, September 1995). Each IRP presents a list of the general outcomes in its subject area; each general outcome is broken down into clearly defined specific outcomes, and accompanied by suggested instructional and assessment strategies and a list of recommended learning resources for teachers.

For example, in the *Mathematics K-7 IRP* one of the K-12 prescribed learning outcomes in the area of measurement is: *describe and compare real-world phenomena using either direct or indirect measurement.* At the K-1 level this becomes *estimate, measure, and compare measures using whole numbers and non-standard units of measurement.* This is broken down into nine specific outcomes, including *use comparative terms to describe time and temperature* and *recognize and name the value of pennies, nickels, and dimes* (Ministry of Education 1995a).

Mary Nall, Field Services Team Coordinator for the Ministry of Education, emphasized that the suggested instructional and assessment strategies that accompany these outcomes are suggestions only. While it is mandatory that the prescribed learning outcomes be met, "teachers need to use their professional judgment and their knowledge of students in choosing the appropriate instructional strategies."

Developing Consensus Among Educators

Consensus among educators on the exact meaning of assessment criteria is as important as the criteria themselves. While clear, unambiguous definitions can decrease the likelihood of misinterpretation, individual teachers will inevitably interpret those definitions through the filter of their own personalities, educational beliefs, and professional experiences. Teachers need opportunities to discuss criteria with colleagues and develop "a common language, common definitions," said Anne-Marie Spizzuoco, a teacher at Morse Street School in Freeport, Maine. "Even the simplest things" can be understood differently, she added. "And then you go to a meeting and say, 'That's not how I did it. Oh, I just assumed'."

Involving teachers in the process of establishing outcomes and defining criteria will facilitate consensus. Morse Street School Principal Cheryl White believes this process cannot be overemphasized. She is reluctant to give out copies of her school's assessment documents, though she is willing to present workshops on how to create them, because she feels they mean little by themselves.

> It's all in the shared meaning, that's where the key is. A lot of people write me and say, "Can we please have a copy of your checklist?" And I say, "The point isn't the checklist! The point is the *process the faculty went through* to beat it out." And that has to be done individually, with every faculty.

Agreement on the meaning and application of criteria is particularly crucial for high-stakes assessments. Herman and her colleagues describe rater training and monitoring procedures that help ensure consistent, reliable scoring of large-scale performance assessments. Training includes extensive discussion of the scoring criteria, practice scoring of sample papers, and decisions about how to handle unanticipated problems involved in scoring a particular set of papers.

Individual teachers or teams can also check how consistently they are applying assessment criteria. NCRESST (undated b) suggests that teachers ask a colleague to score a set of papers or portfolios, or rescore items themselves after an interval of two or more months, and compare the scores.

The Tools of Authentic Assessment

Some of the tools of authentic assessment will be familiar to longtime teachers. Others may be new and strange. No single method is sufficient by itself, nor must every teacher use every method. Different teachers will select different tools from the assessment "toolbox" and adapt them to their own needs. What is important is that a variety of methods should be used to gather evidence of student learning, in a variety of settings over time. This way methods with different strengths and weaknesses can complement each other, and fluctuations in student performance at particular times or in specific settings will not carry inordinate weight in the total assessment picture.

Observation is one of the mainstays of assessment in developmentally appropriate programs. "Observation of children is the most significant way in which the teacher learns about the children, how they learn and how they make sense of the world," emphasize the creators of the British Columbia Primary Program (Ministry of Education 1990c). Many of the tools of authentic assessment are designed to help teachers record, organize, and make sense of their classroom observations.

Anecdotal Records

Anecdotal records are informal written notes of teachers' observations of student behaviors, actions, and reactions during the course of normal classroom activities, "milestones particular to that child's social, emotional, physical, aesthetic, and cognitive development" (Thomas C. Boysen 1993). They can be kept using many different means, from sticky notes or index cards carried on a clipboard during the day and transferred to students' files at the end of the day, to preprinted charts with students' names listed along one side and columns for recording observations of behavior in different learning areas.

Anecdotal records should be objective, judgment-free descriptions of behavior—raw assessment data, so to speak, with analysis and interpretation of those data reserved for a later time. They focus on the positive rather than on the negative: on what a child does, rather than on what he or she *does not* do or *should* be doing. "Behavior is what you see. Processes and motivations are not seen," Jim Grant and Bob Johnson (1994) explain. "You may infer these, but first record the behavior free of any interpretation or judgment. This allows you to collect a pattern of behavior and over time you may find you change or alter your interpretation."

For example, "D checked out 4 bear books from the library" describes behavior, while "D loves bears" and "D wouldn't share any bear books with other children. He hogged them all" are judgments. "S ran around the kick ball field for about 3 minutes when I asked for the students to line up" is an observation, but "S does not listen" and "S shows off for other children by disobeying rules" make assumptions about motivations (Thomas C. Boysen 1994).

Focussing on behavior can reduce the effect of unconscious biases on observer perceptions. One such bias, the *halo effect*, is the tendency to notice and remember more negative behaviors if one has negative feelings about a child, and more positive behaviors if one has a generally positive opinion of the child. Another, the *logical error*, is the tendency to assume an unobserved behavior is present because it is often associated with an observed behavior—for example, rating a child's comprehension higher than it actually is because the child often chooses to read during choice time (Grant and Johnson).

Teachers should not be dismayed if they find anecdotal records difficult and time-consuming at first. As David Elkind emphasizes, "Observation is a skill that has to be learned and is not something that all teachers can do without training." The education expert recalls that it took him years to learn how to interpret children's drawings and other works (Ministry of Education 1992). Like any new skill, taking anecdotal records becomes easier and more automatic with time, training, and practice.

The teachers at Bronx New School initially felt overwhelmed by the time demands of anecdotal records, but ultimately decided they were well worth the effort.

> Through their experiences of writing down their observations in a variety of settings and a variety of ways,most learned so much more about their students that they eventually became staunch advocates of keeping written records. They saw that memory of the details and the nuances... does indeed escape them in the blur of time passed; that only by writing down observations can teachers achieve a perspective of each student's unique growth over time. (Falk)

It should be noted that the informal anecdotal notes teachers record for their own use are not the same as the formal anecdotal progress reports described in chapter 3.

Rating Scales

Rating scales can be helpful guides to informal observation. In addition to identifying specific behaviors to look for, they often provide an organized structure that places those behaviors in a developmental or theoretical context.

The *checklist* is the simplest form of rating scale. Checklists contain brief descriptions of dimensions, characteristics, or behaviors; the list of descriptions may be long, but each description is usually quite concrete and specific. The teacher using the checklist simply indicates the presence or absence of each behavior or dimension (Herman and others).

True *rating scales* resemble checklists, but rather than simply indicating "yes" or "no," the teacher notes the extent to which the behavior is in place or how well the task was accomplished. Rating scales can be numerical or qualitative. "A numerical scale uses numbers or assigns points to a continuum of performance levels. The length of the continuum or the number of scale points can vary.... A qualitative scale uses adjectives rather than numbers to characterize student performance" (Herman and others).

For example, Lescher presents a "Primary Reading/Writing Checklist" that lists behaviors such as "Displays understanding of letter-sound association" and "Recognizes some words by sight." The teacher rates each item of behavior U for Usually, S for Sometimes, or N for Not Yet. While it is called a checklist, this is actually a three-point qualitative scale. Rating scales vary in complexity and sometimes contain both numerical and descriptive elements.

Rubrics are sets of criteria designed to score complex performance tasks. "A typical rubric states all the dimensions being assessed, contains a scale, and helps the rater place the given work properly on the scale" (NCRESST undated a). For example, Fairplay Elementary School's *Quest*

performance assessment is scored with several rubrics that use a 6-point scale. A score of 1 on the *performance* rubric indicates, in part, that the oral presentation "Lacks imagination and originality" and that "Evidence of risk-taking and resourcefulness is missing," while a score of 6 indicates that it "Displays exceptional imagination and originality" and "Shows a great deal of risk-taking and resourcefulness" (Julie McCann undated).

Checklists, inventories, and rubrics can aid teachers moving into the new territory of authentic assessment. But they can also overwhelm novice innovators with more information than they can absorb or use effectively. "Checklists can be hazardous to your health if you try to implement them all at once!" warns Janine Batzle (1992). Grant and Johnson suggest that teachers examine a number of existing checklists and inventories, then draw on those models to create their own. "The process of creating your own checklists or inventories will help you become conversant with developmental patterns and provide a good background for your observations. Keeping inventories on every child can become overwhelming and may not be the best use of your time. But having worked through the inventories and having them for reference will give you a firm base" (Grant and Johnson).

Systematic Observational Assessments

Informal classroom observation can yield valuable insights into children's learning processes, interests, and unique perceptions. Systematic or standardized observations have the additional benefit of allowing reliable comparisons among children or between one child's performances at different times. The best known systematic observational assessments may be those developed by New Zealand educator and psychologist Marie M. Clay, the founder of the world-renowned, highly successful Reading Recovery tutoring program. These assessments include *running records*, which have been shown to have reliabilities as high as .90, and a battery of other, complementary literacy assessments designed for beginning readers (Marie M. Clay 1993).

In these systematic observations, a child is observed while engaged in an authentic, yet standardized task, administered in a standard way; observer training and assessment procedures are designed to reduce observer bias. For example, in taking a running record a teacher listens to a child read aloud a one to two hundred word text and records on a score sheet everything the child does, including correct responses, types of errors, and self-correcting behaviors (Clay).

Running records can reveal a wealth of previously unnoticed information that more than justifies the time and practice required to master them (Batzle). Running records are especially helpful in revealing evidence of reading progress too subtle to be measured by standardized tests (Falk).

When she developed the technique, Clay herself was surprised to discover how much new information it yielded:

> I had been teaching reading and remedial reading for many years when I began my research on emergent reading behavior. I am still humble about the fact that I had never noticed self-correction behaviour until I started recording everything that children were doing. It was then I found that I had been missing something that was very important.

Clay recommends that teachers receive training in systematic observation techniques, or at least have opportunities to discuss them with colleagues, rather than trying to implement them solely on the basis of reading her books.

Tests

Paper-and-pencil tests have a legitimate place in the multiage teacher's assessment repertoire. It is only exclusive reliance on them that is problematic. Paper-and-pencil tests borrowed from an old "scope and sequence" math text helped rescue Terry Snyder from the reporting dilemma described in this Bulletin's introduction. Although conventional tests could not measure the enjoyment of math Snyder saw developing in his mixed-ability class, they did provide reassuring documentation that his students were mastering basic skills and concepts at least as well as the students he had previously taught in homogeneous math groups.

Westmoreland primary teachers also use pre- and post-tests to assess math and literacy skills at the beginning and end of each school year, explained Snyder's colleague Carol Olson. With assistance from district assessment staff, the team borrowed elements from several existing tests, including an old district reading-skills test, to create a relatively short but comprehensive test that helps them ascertain where their students are starting from in the fall and how much progress each one has made by the following spring. Pretests for younger students use unwritten tasks such as identifying letter names and sounds. Results of the past two years' posttests showed that most of Snyder's second graders met 95 percent of the second-grade goals, said Snyder, with 86 percent being the lowest score.

Conversations with Children

Conversations with children help teachers understand children's attitudes toward themselves and school. The British Columbia Ministry of Education considers conferences and conversations with children one of the most important ways of collecting Primary Program assessment evidence (1990b). The Ministry suggests teachers schedule a formal, individual con-

ference with each child once a week in addition to the informal conversations that occur during normal classroom activities (Ministry of Education 1990c). It suggests teachers keep conference logs and provides a "Child-Teacher Conference Sheet" with suggested questions to ask the child.

Child-Teacher Conversations are also an important part of assessment in the Kentucky Primary Program. The KELP Teacher's Handbook provides a list of suggested questions that may be used to initiate conversation, but stresses that the teacher's role is to be "an active and supportive listener" (Boysen 1994). Teachers are asked to read their notes back to the child at the end of the conversation and ask if he or she has any comments to add before signing the conversation record.

Structured oral interviews and reading, writing, and attitude surveys can also help teachers discover students' perceptions of the learning process (Batzle, Lynn K. Rhodes 1993, Boysen 1994).

Portfolios

Portfolios are a means of collecting, organizing, and reviewing authentic-assessment evidence gathered over time. Sometimes the term *portfolio assessment* is used to indicate the whole range of authentic assessment practices described in this Bulletin. The following definition was created by educators from seven states participating in a 1990 conference sponsored by the Northwest Evaluation Association:

> A portfolio is a purposeful collection of student work that exhibits the students' efforts, progress, and achievements in one or more areas. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit, and evidence of student self-reflection." (NCES 1995d)

Portfolios contain samples of student work ranging from the traditional drawings, math papers, and writing samples to audio and videotapes with recordings of student presentations or oral assessments. Portfolios may also include teacher observations, results of standardized tests and other formal assessments, notes on student-teacher conferences, parent surveys and comments, and peer evaluations. "Portfolios offer a wonderful visual presentation of a student's capabilities, strengths, weaknesses, accomplishments and progress," comments Batzle.

Three of the most common types of portfolio are the *working portfolio*, the *showcase portfolio*, and the *record-keeping portfolio*. The showcase portfolio, which is a collection of "best work," is under the total control of the student. The prospect of collecting and proudly displaying their creations can powerfully motivate children. However, as Batzle points out, best pieces alone do not give a wholly accurate picture of a students' work and abilities.

BULLETIN IN BRIEF

OREGON SCHOOL STUDY COUNCIL

Condensed from OSSC Bulletin | Vol. 39, Nos. 3 and 4 • January and February 1996

Assessment and Evaluation in the Multiage Classroom

By Joan Gaustad

Assessment of student progress is a challenge for educators who use developmentally appropriate practices such as multiage grouping. Assessments that were designed to work with conventional age-graded practices may not work as well in the multiage clasroom. Even many educators who work in the conventional age-graded classroom are dissatisfied with the limitations of conventional assessments.

Interest in alternative types of assessment has thus become widespread. These performance-based, or *authentic*, assessments are being vigorously discussed, researched, and implemented across the continent.

What Is Good Assessment?

Each type of assessment has strengths and weaknesses; no single assessment method serves all needs equally well. Which characteristics are most important depends on the purpose the assessment is to serve and the audience that will ultimately use the information it produces.

Conventional Assessments

Most adults recall these methods from their elementary-school days: paper-and-pencil fill-in-the-blank, and multiple choice tests; letter-grade evaluation; grading on a "curve"; and computer-scored standardized tests. These assessments began to come into use early in this century and became more common as the decades passed. Over time, considerable evidence has accumulated concerning their negative effects on student learning.

Conventional classroom assessment and evaluation methods tend to be subjective and inconsistent. High-quality training in assessment and reporting might reduce these problems, but few teachers and fewer administrators have had such training. Competitive grading destroys collaboration and community. It damages the self-esteem of slower learners. And it can negatively affect the motivation and quality of work of both able and less-able students. In any case, grading on a curve is statistically invalid for the small number of students in one school or class.

Well-designed standardized tests are efficient, economical, and far more statis ically valid and reliable than teacher-created tests. However, test users have seldom understood their limitations. Standardized tests have often been misused, overused, and misinterpreted, and questionable results have been used to make significant decisions. Critics also charge that test questions are biased against students of different cultural, linguistic, and socioeconomic backgrounds.

Neither standardized tests nor conventional classroom assessments measure the most important aspects of learning. Conventional assessments are based on outdated assumptions concerning the learning process. They focus on lower order, decontextualized skills and do not assess

The Oregon School Study Council—an organization of member school districts in the state—is a service of the College of Education, University of Oregon.

This issue of Bulletin in Brief is condensed from *Assessment and Evaluation in the Multiage Classroom*, by Joan Gaustad, OSSC Bulletin, Vol. 39, Nos. 3 and 4, January and February 1996. 70 pages. For ordering information, see page 4.

higher order skills or the ability to apply knowledge in context. An "overwhelming body of educational research" shows that students with good grades and high standardized test scores typically don't understand what they have studied (Howard Gardner 1991).

Authentic Assessments

Rather than allowing efficiency and economy to determine assessment design, the developers of authentic assessments set ambitious goals:

1. To assess truly important aspects of learning
2. To assess learning comprehensively
3. To be responsive to the needs of individual learners
4. To positively affect instruction
5. To gain both validity and reliability
6. To attain practical feasibility

It is too soon to know how fully these goals will ultimately be achieved. Authentic assessment currently appears most successful in the goal areas where the shortcomings of standardized testing are most evident, and weakest in the areas where standardized testing is strongest: efficiency, economy, and statistical validity and reliability.

A national panel on assessment reform recently suggested that large-scale, high-stakes performance-based testing be postponed until such assessments have been refined. However, the panel recommended the immediate adoption of alternative assessments at the classroom level. "To wait until fully developed standards and assessment instruments are available in all knowledge domains, and capacity to implement them exists in all areas of the country, as opposed to acting on what we know now, would cheat many American students out of beneficial changes," said the panel's report (Karen Diegmueller 1995).

Assessing Student Learning: Products and Process

In the conventional classroom, assessment focuses primarily on the products of the learning process: student writings, drawings, test papers, and so forth. In the multiage classroom, *processes* of student growth and learning are also assessed.

Setting Clear, Assessable Objectives

Authentic assessment considers children's growth and development in social, emotional, and physical areas as well as in intellectual areas. Before progress can be assessed in any of these areas, clearly stated, measurable objectives must be established. As many stakeholders as possible should be involved in defining major instructional goals.

Next, specific performance criteria for meeting those objectives must be agreed upon by educators and communicated to students and parents. Once objectives and criteria have been set, instruction can be planned and appropriate assessment tools selected to monitor progress toward them.

The Tools of Authentic Assessment

Some authentic-assessment methods will be familiar to longtime teachers, while others are new. Different teachers will select different tools from the assessment "toolbox" and adapt them to their own needs. The key is that a variety of methods should be used to gather evidence of student learning in a variety of settings over time.

Observation, one of the mainstays of assessment in developmentally appropriate programs, can yield valuable insights into children's learning. Many of the tools of authentic assessment are designed to help teachers record, organize, and make sense of their classroom observations.

Anecdotal records are informal written notes of teacher observations of student behaviors, actions, and reactions during the course of normal classroom activities. They can be kept in a variety of ways, but should be objective, judgment-free descriptions of behavior—raw assessment data, so to speak, with analysis and interpretation of that data reserved for later.

Rating scales help guide informal observation by identifying specific behaviors to look for, and often provide an organized structure that places behaviors in a developmental or theoretical context. The simplest type of rating scale, the checklist, contains brief descriptions of specific characteristics or behaviors; the teacher simply indicates the presence or absence of each behavior. With more complex rating scales, the teacher notes the extent to which the behavior is in place

or how well the task was accomplished.

Systematic observational assessments such as the *running records* developed by Marie M. Clay (1993) allow reliable comparisons among children or between one child's performances at different times. Children are observed while engaged in an authentic, yet standardized, task, administered in a standard way; observer training and assessment procedures are designed to reduce bias.

Paper-and-pencil tests have a legitimate place in the multiage teacher's assessment repertoire. It is only exclusive reliance on them that is problematic.

Conversations with children help teachers understand children's attitudes toward themselves and school. Such conversations are important components of assessment in the Kentucky and British Columbia primary programs.

Portfolios are means of collecting, organizing, and reviewing authentic-assessment evidence. A key element of portfolio use is student participation in the selection process. In addition to student work samples, portfolios may include teacher observations, results of standardized tests and other formal assessments, notes on student-teacher conferences, parent comments, and peer evaluations.

Involving Students in Assessment and Evaluation

Authentic assessment involves students as active participants in establishing learning goals, evaluating their own and their peers' achievements, and reporting progress to their parents. Students are encouraged to reflect on what they are learning and on the learning process: to recognize their accomplishments and areas where they can do better, to understand what learning strategies work well for them, to evaluate and consciously strive to improve their work. Children need considerable modeling and instruction from teachers to begin to develop these important skills.

Reporting Student Progress to Parents

To educators who have been studying and gradually implementing developmentally appropriate practices over the course of several years, changes in reporting may seem like just one more step in a natural, evolutionary process. It is easy to forget how revolutionary multiage practices can appear to adults whose images of elementary school have not changed since their own childhoods. Unless parents have been informed and involved throughout the change process, a new report card may be perceived as a sudden, shocking abandonment of tried-and-true practices. Educating parents about curriculum and assessment is therefore an essential part of the reporting process.

Ongoing Communication with Parents Is Crucial

Parents who understand and support developmentally appropriate practices will be receptive to complementary changes in assessment and reporting. Means of informing parents about innovations include letters, school newsletters, and curriculum information enclosed with report cards. Parent meetings provide opportunities for indepth explanations and exchanges.

Communication with parents should be a two-way street. Parents' special knowledge of their children can be extremely helpful to teachers. Parent surveys and questionnaires can tap this knowledge. Parents can also be invited to observe in the classroom and to participate in some aspects of assessment. Parental feedback should be solicited early in the process of changing progress reports.

Progress Compared to What?

Parents need a context to understand their child's progress. Parents deserve to know how their children are doing in terms of individual growth and progress toward established learning objectives. Reporting progress with reference to exit criteria—criteria that must be met before a student can pass to the next educational level—answers the question "Is a student on course or not to get to the destination on time?" while deemphasizing inappropriate comparisons with classmates' rates of progress.

Parents also want to know how their children's abilities and achievements compare to those of other children. The British Columbia Ministry of Education created sets of "widely

held expectations," developmental continua that describe behaviors and abilities that normally develop within certain age ranges. These provide accurate yet flexible frames of reference for children's development (Ministry of Education 1991).

Written Progress Reports and Conferences Are Complementary

Descriptive comments and rating scales are the chief components of most multiage written progress reports. However, even the best written reports can only communicate bare-bones information about a student's progress. In-person conferences put meat on those bones and provide opportunities to build rapport with parents, educate them about instructional and assessment practices, prevent or clear up misconceptions, and obtain information about the student and the family context.

Children should be involved in the conferencing process. Teachers may meet with students before the parent conference, or students may participate in three-way conferences with parents.

IMPLICATIONS FOR ADMINISTRATORS AND SCHOOL BOARDS

Administrators and school board members have two main roles to play in relation to authentic assessment: to support teachers, and to promote understanding and acceptance of developmentally appropriate assessment and reporting practices in the community.

To be successful, authentic assessment must be implemented as part of comprehensive changes in curriculum and instruction. Teachers need considerable staff development and long-term technical and psychological support to make these changes. They also need extra planning time on a permanent basis, as authentic assessments are more time-consuming than conventional methods. Providing this time and support will require extra financial resources. Principals should ideally have excellent interpersonal, management, and fund-raising skills as well as understanding of, and commitment to, authentic assessment.

Many administrators and school board members will need to make themselves more assessment-literate before they can begin educating the community. They cannot be effective spokespersons for authentic assessment unless they understand its potential and the shortcomings of conventional methods. Their goal should be to become open-minded yet critical consumers of all methods, remembering that authentic assessment is still in its infancy despite its great potential.

Bibliography

Clay, Marie. *An Observation Survey of Early Literacy Achievement.* Portsmouth, New Hampshire: Heinemann, 1993. 93 pages. ED 358 427.

Diegmueller, Karen. "Push Ahead With Standards, Alternative Assessments, Panel Urges." *Education Week* XIV, 41 (August 2, 1995): 6-7.

Gardner, Howard. *The Unschooled Mind: How Children Think and How Schools Should Teach.* New York, New York: Basic Books, 1991. 303 pages. ED 368 451.

Ministry of Education, Province of British Columbia. *Supporting Learning: Understanding and Assessing the Progress of Children in the Primary Program. A Resource for Parents and Teachers.* Victoria, Province of British Columbia, Canada: Author, 1991. 62 pages.

ORDER INFORMATION

Additional Copies of this Bulletin in Brief can be ordered at the rates below:

	Member Price	Nonmember Price
1-9 copies	$1.50	$2.00
10-24 copies	1.25	1.65
25-49	1.00	1.35
50-99	.75	1.05
100+ copies	.50	.75

All rights reserved. This publication may not be reproduced without permission in writing from the publisher.

Copies of the *OSSC Bulletin* from which this *Bulletin in Brief* has been condensed are also available, at $10.00 per copy.

Full payment or purchase order must accompany all orders. Make checks payable to **University of Oregon/OSSC.** Shipping is by UPS Ground or equivalent. For faster delivery, call for charges. Order from: **Oregon School Study Council, 5207 University of Oregon, Eugene, Oregon 97403-5207.** (800) 438-8841 or (541) 346-5044. Fax (541) 346-2334.

Showcase portfolios are not very useful for guiding instruction and may give parents a one-sided perception of their child's development, exaggerating real strengths while deemphasizing equally real weaknesses. Teachers whose students create showcase portfolios may also develop a record-keeping or teacher portfolio in which additional samples and other district records are kept (Batzle).

If unnecessary papers are not periodically weeded out, a portfolio will end up resembling "a kitchen gadget drawer, so chock-a-block with unrelated items that locating the corkscrew becomes a frustrating quest" (Maeroff). Aides or volunteers can help with such vital details as labeling and date-stamping student work. But the teacher must be responsible for selecting key pieces worth retaining, or for teaching students how to evaluate their work and make such selections themselves. "It's a sloppy, time-consuming job," said Snyder. "And you don't know exactly what's going to be monumental for a child and what's not; you don't know which day the breakthrough is going to come. So you have to spend time flipping through individual papers." Nonetheless, Snyder feels the information gained from portfolios justifies the effort. "A report card is a snapshot that doesn't mean much by itself. A portfolio with examples of a child's work over time is more like a video," he said.

Involving Students in Assessment and Evaluation

Authentic assessment seeks to involve students as active participants, not just as passive objects. Students may participate in all stages of the process, from establishing group or individual learning goals to evaluating their own and their peers' achievements, to reporting their progress to their parents.

Multiage teaching encourages students to consciously reflect on what they are learning and on the learning process: to recognize their accomplishments and areas where they can do better, to understand what learning strategies work effectively for them, to evaluate and consciously strive to improve their work. Work samples of consistently improving quality are less important for their own sake than as evidence a student is developing these important abilities.

Communicating Objectives and Criteria to Students

Involving students in assessment begins with communicating to them the objectives they are expected to achieve and the criteria by which they will be judged. Fairplay Elementary School faculty began this process by putting its outcomes into "student language" and posting them in each classroom, said Principal Julie McCann.

It is also helpful to show students samples of work that meet the criteria to use as models. Wolfe and her colleagues present examples of successful and unsuccessful third- and fifth-grade student writing for use in Oregon classrooms. They suggest ways teachers can use the examples to deepen their student's understanding of quality writing, including reading the samples aloud, discussing the differences between strong and weak writing samples, and teaching students to use the state criteria to score writing samples. Also included are written "guides to revision" designed to stimulate students to examine their own compositions and think about how well they meet the criteria.

Students, like teachers, are more likely to understand and remember objectives and criteria they participate in creating. For example, students in Gretchyn Turpen's elementary class came up with the following list of criteria for a "best" piece of writing: "Uses imagination; The words make pictures in your mind; Uses things that happen in your own life; Uses writing rules like periods and capitals; Spaces between words; Makes you want to keep on reading" (Bridge and others). Individual students can also create lists of objectives they want to work toward, such as improving punctuation or character development. Then they can evaluate their finished work against the list and check off the objectives they think they have met (Allan De Fina 1992).

Student Self-Evaluation

Self-evaluation can be difficult even for adults. Children need considerable assistance from teachers to begin to master this complex, important process.

Portfolios offer teachers many opportunities to guide and model self-evaluation. Each time they confer with their students about which items to retain is another opportunity to connect concrete examples to abstract criteria. Asking students to make decisions about the quality of their work, comparing one piece to another, helps them comprehend and internalize evaluative standards, says De Fina. Students "will soon begin to recognize that some work samples do not demonstrate certain desired qualities, and since they naturally want to show their best work, they will begin to carefully choose items for inclusion or showcasing" (De Fina).

Until students have considerable facility with self-assessment, it is risky to give them full responsibility for high-stakes assessment choices. Maeroff cites the unfortunate example of a Rhode Island third-grader who kept displaying examples from the beginning of the school year when she was asked to show her "best" work in a statewide assessment, though her work had presumably improved during the term.

Penelle Chase and Jane Doan (1994) used portfolios solely for "low-stakes" purposes. Believing their evaluation system was already working well, they decided to explore portfolios as a growth experience for their five- to eight-year-old students rather than as a significant source of assessment data. The two coteachers acted as models by sharing their own professional portfolios, then let their students take the lead in the enthusiastic discussion concerning what student portfolios should include. "Knowing that we did not ourselves need the portfolios for evaluation purposes, we were happy to allow the children to assume control of the contents of portfolios," Chase comments.

While no conclusive evidence yet exists, interviews with fourth-graders who participated in the 1992 NAEP reading assessment found indications that self-evaluative skills were related to reading proficiency. Students who had selected their own writing samples to bring to the interview had higher average reading performances than those whose teachers had made the selections, and students who had some ability to evaluate their classroom work samples scored higher than those who could not (NCES 1995a).

Peer Evaluation

Chase and Doan describe many ways they help students learn to evaluate their own and others' work. They model and encourage positive rather than negative feedback in group discussion, challenging children to be specific in their praise: "Why do you like the way Hannah wrote it?" or "What words that Tracy used were good words?" Once each day, student and teacher observers watch classroom interactions, recording praiseworthy social behaviors and effective learning strategies, and later reading their notes aloud to the class. These practices reward positive accomplishments, highlight successful strategies, and encourage students to reflect on the learning process and learn from each other as well as from the teachers.

The desire for peer approval can motivate children to improve their work. For example, Turpen's students were inspired to do their best by the prospect of presenting their writings to an audience. "Now when the authors read their best pieces in the classroom, they aren't satisfied unless they are getting laughs, gasps, or smiles from the audience," she reported (Bridge and others).

Shepard describes how third-grade teachers involved in one assessment research project used scoring criteria to teach their students to write better story summaries. At first students were undiscriminating, scoring all their classmates' summaries 3 on a 4-point scale. One teacher wrote bad summaries for the children to score, compared them to summaries on book jackets, and used this as basis for class discussion. Other teachers had whole

classes read a story and develop a summary as a group, with much discussion along the way. "Eventually students got much better at writing summaries as a result of their teachers' effective use of modeling and class discussions about using scoring rubrics" (Shepard 1995).

Fairplay students use rubrics to score their own and classmates' work, and, according to McCann, many have come to realize the value of feedback and to seek it out in order to improve. "If someone gives every paper top scores, the kids won't accept that. They'll say, 'You didn't spend time assessing my paper. I know this is not all sixes'."

Fairplay Students Create Their Own Assessment Challenges

Student involvement in assessment is one key to the success of Fairplay Elementary School's multiage program. When the Oregon educational reform act passed in fall 1991, Fairplay staff had already begun implementing multiage practices and were rethinking their approach to assessment. The following year the Corvallis school was one of seven Oregon schools funded to develop performance assessments to judge student progress toward the proposed tenth-grade *Certificate of Initial Mastery* outcomes (Oregon Department of Education, Fall 1992). "We didn't have a lot of expertise and knowledge," said Principal McCann, "but we had some real, legitimate questions we wanted to answer."

A particularly successful innovation was the schoolwide performance assessment called the Quest. All third, fourth, and fifth graders (the program retains traditional grade labels for convenience) are required to create their own assessment task: to formulate a question whose answer they do not know in a general topic area and, within a two-week period, to research and prepare visual, oral, and written presentations answering the question. Each dimension of the project is scored according to a rubric with a 6-point scale, copies of which are given to students in advance to guide their work. Fifth-graders must present their Quest projects to panels of educators, parents, and community members (McCann undated).

The individualized challenge of the Quest fired the imaginations of Fairplay's students and became a highly successful motivator, said McCann.

> We began to see many children's attitudes shift... they were excited about what was going on and wanting to know more. One student finished his Quest in two days instead of the allotted two weeks, and he was complaining about how bored he was and how mad he was. His teacher said, "Well, so what will you learn from this?" And he replied, "To come up with a harder question next time," which is very different from "I just want to get this done."

Fairplay's multiage groupings allow younger children to observe and

be inspired by the example of their older classmates. "We saw first and second graders already planning what their Quest was going to be," related McCann. "They began to make connections: "I need to know how to write a paragraph for my Quest, I need to know how to use correct grammar, I need to know how to speak in front of an audience."

The effects of Fairplay's changes in assessment and instruction are beginning to show up in state test scores. Before restructuring, Fairplay typically ranked ninth or tenth out of the district's ten elementary schools in state test scores; now it is in the top five. And in Oregon's first open-ended math assessment, given in spring 1995, Fairplay fourth graders scored highest in the district (Fairplay Elementary School undated).

Conclusion

Authentic assessment methods can only be implemented successfully if simultaneous changes are made in curriculum and instruction. But a developmentally appropriate multiage program cannot be implemented overnight. "You will fail if you expect instant results," cautioned McCann, noting that Fairplay actually experienced decreased scores in some areas in the first years of change. "We've been at it six years and I don't think we've peaked yet," she added.

Even if educators are willing to make a long-term commitment to change, to invest time mastering new instructional and assessment approaches, and to engage in a ongoing process of discussion and mutual learning with colleagues, major problems can surface at reporting time if assessment changes take parents by surprise. How to prevent or minimize such problems is the topic of chapter 3.

Chapter 3

Reporting Student Progress to Parents

To teachers and administrators who have studied developmentally appropriate practices, undergone extensive staff development, and gradually implemented new methods over the course of several years, changes in reporting may seem like just one more step in a natural, evolutionary process. It is easy to forget how revolutionary multiage practices can appear to adults whose images of elementary school were formed in their own childhoods and have remained unchanged since then.

Unless parents have been informed and involved throughout the change process, a new report card may be perceived as a sudden, shocking abandonment of tried-and-true practices. In Kentucky, where education reform mandated revolutionary changes in all aspects of primary education, parents "seemed more anxious about changes in the report card" than about anything else (James Raths and others 1992). Educating parents about curriculum and assessment is an essential part of the reporting process.

What! No Grades?

Many parents are unaware of the inconsistent and subjective nature of letter grades (described in chapter 1). Like Jean Germani, a Rhode Island parent fighting against elementary school reporting changes, they believe ABC grading is "universally understood and tells parents 'exactly, precisely, where our children stand. It's so objective'" (Lynn Olson 1995).

In addition, Americans' "deep-seated national faith in competition" (Anderson and Pavan) makes it hard for many people to accept that classroom competition has drawbacks. Some parents of high-achieving students are so eager to obtain proof of their children's superiority that they are

reluctant to consider the negative effects of ranking. Others are convinced students will lose motivation without the prodding of grades. Even parents who fear their children will be unsuccessful, and whose personal memories of school are unpleasant, may be uneasy about abandoning the familiar evil of competitive grading for unfamiliar methods (Falk).

Parents of primary-age students may be more receptive than parents of older students to reporting methods compatible with authentic assessment. In many primary schools, ABC grades have traditionally not been given to the youngest children. Carol Olson recalls that when she was a student in the Eugene School District, she did not encounter letter grades until middle school, and she has not seen them used at the elementary level during her teaching career with the district. But even so, a few Westmoreland parents called for ABC grades when they were surveyed about report-card changes last year, she said.

Parents, like educators, need time to learn about new instructional and assessment methods and judge their worth. Parents deserve to be treated as intelligent partners in their children's education. It is up to educators to show parents that authentic assessment can give them more and better information about their children's learning than conventional methods.

Ongoing Communication with Parents Is Crucial

Communication with parents is always important, but it is particularly crucial when major changes are occurring. Happily, multiage classes in which children remain with teachers for more than one year provide additional opportunities to establish good communication with parents. Parents who understand and support developmentally appropriate practices will be receptive to complementary changes in assessment and reporting.

Giving Parents Information

Educators can begin informing parents about innovations via traditional methods of communication such as letters to parents and school newsletters. When Westmoreland primary teachers began exploring multiage methods, they simply "enhanced" existing avenues of communication to get the word out to parents, said Olson (Gaustad 1992a). Chase and Doan present a sample letter to parents that embeds explanation of whole language and thematic instruction in a cheerful, chatty update on current class activities. In programs using portfolios, students may be asked to take work samples home and bring them back with parental comments in writing (Batzle, Lescher). Informational materials such as term outlines and curriculum overviews can also be enclosed with report cards (Ministry of Education 1990c).

Parent meetings provide opportunities for indepth explanations and exchanges. Fairplay Elementary School held a series of meetings that focused on whole language, manipulative math, and other topics requested by parents. Fairplay also scheduled activity nights during which kindergartners and their parents engaged in a series of process-writing activities. Then the children left the room and their parents discussed the experience with Fairplay's principal and kindergarten teacher. "We shared with them how this method compared to the way we were taught as children and began to point out the similarities. We talked about things to look for in children's writing so that they understood the developmental process as it was unfolding with their child" (Fairplay Elementary School 1994).

Morse Street School Principal Cheryl White augmented in-person parent meetings with a cable television presentation to spread the word about the school's new assessment system.

Seeking Information from Parents

Communication with parents should be a two-way street. Parents' special knowledge of their children can be extremely helpful to teachers. The Westmoreland primary team seeks to tap this knowledge by sending parents of new students a checklist and questionnaire before their first conference of the year, which takes place about six weeks into the fall term. The "Menu of Student Characteristics" asks parents to rate their children on items such as "needs lots of attention and praise," "works well independently, and "likes math challenges." The questionnaire includes prompts such as "What would you like your child to learn this year?" and "My child could spend a whole day doing...." It also asks for a basic health history and inquires about recent family problems that can significantly affect children, such as death, divorce, or job loss (Westmoreland Elementary School undated a). Teacher and parents discuss this information along with teacher perspectives as they decide on the goals the child should work toward during the year, Snyder explained.

Anthony D. Fredericks and Timothy D. Rasinski (1990) suggest several ways to involve parents in assessing their children's reading. They provide an attitudinal scale and observation guide parents can use at home to assess their children's reading skills and attitudes. "Parents' participation in evaluating their children's growth can help to eliminate many misconceptions and misinterpretations that may occur during parent-teacher conferences or at report card time," the authors comment.

Educators can also invite parents to observe multiage instruction in action. The British Columbia Ministry of Education (1990c) has created several forms designed to guide parents in observing their child's learning and behavior in the multiage primary classroom.

Communicating About Reporting

Since the purpose of a progress report is to communicate to parents, parents must be the final judges of how effectively it does its job, says Grant Wiggins (1994). Soliciting parent feedback early in the process may help educators reduce later revisions, along with building trust and goodwill. British Columbia Principal Trevor Calkins (1992) reports that the openness of the decision-making process and "the feeling that the decision was not made ahead of time" were two reasons parents gave for their willingness to accept an anecdotal reporting system.

The staff of Clyde Miller Elementary in Aurora, Colorado, involved parents from the beginning when they decided to change reporting to reflect their authentic-assessment practices. Parents who had participated in developing the performance-based report card explained it to the rest of the parents at a special PTA meeting, assisted by students and staff. The first time the new report card was sent home, teachers included a brochure explaining how it differed from the old card and a survey asking for comments and questions. The card was then modified in light of parent, student, and staff suggestions (Evelyn Kenney and Suzanne Perry 1994).

The amount of effort necessary to inform parents will vary from community to community, observed Morse Street School teacher Spizzuoco. "Our parents are very involved. We have almost one hundred percent attendance at parent conferences. So not many were surprised about reporting changes. But if you have a system where parents are less involved, then it's going to take more parent education. You have to know your system to know what it will take to get parents on board."

Progress Compared to What?

"To know how a child is doing, the parents need a context: compared to what?" Wiggins explains (1994). Two key concepts—*progress* and *growth*—must be differentiated in establishing a reporting context: "Progress is measured backwards from the goal, and growth is typically defined as change in the student. But a student could change a great deal without making much progress toward the standard" (Wiggins 1994). A third alternative, normative comparisons, uses the context of other children's rates of progress.

Parents deserve to know how their children are doing in terms of both growth and progress. Parents also want to know how their children's abilities and achievements compare to those of other children. The challenge is to give parents the comparative information they need to answer legitimate questions, while minimizing damaging and misleading comparisons among classmates.

Individual Growth

According to Anderson and Pavan, the most important types of information teachers can give parents are (1) a complete, accurate description of their child's potential, and (2) an evaluation of the extent to which their child's performance fulfills that potential. The healthiest form of competition is competition with oneself, and all children, whatever their capabilities, should be challenged to improve over their own previous performances.

However, considering the erratic and individual nature of development in young children, Goodlad and Anderson urge educators not to make premature judgments about children's academic, social, physical, or emotional potential, and to be cautious about sharing predictions with parents. Such judgments can create lowered expectations that become self-fulfilling prophecies. Descriptive statements should therefore make up the preponderance of reporting comments in the primary years. More evaluative comments may be included as students grow older.

General Normative Comparisons

Most parents appreciate receiving detailed information about their child's individual strengths and weaknesses. But inevitably they also ask, "How is my child doing in relation to other children his/her age?" (Falk). This question masks two main concerns, said Merry Denny: "Do I need to be worried? and "Is my child gifted?" The answers to these questions affect how parents plan for their children's future.

To meet teachers' and parents' needs for comparative information, the British Columbia Ministry of Education created sets of *widely held expectations* for each of the five areas in which it sets goals and assesses student progress, as well as for development in reading, writing, and mathematics (Ministry of Education 1991). These developmental continua describe behaviors and abilities that normally develop within certain age ranges, providing accurate, yet flexible frames of reference to help parents understand their children's development. For example, in the area of physical development, children from three to five years in age "begin to understand and use concepts of place and direction—up, down, under, beside." Children seven to nine years in age "continue to develop an understanding of direction and place although many confuse right and left, up and down when playing games." Between the ages of nine and eleven children "are developing ability to coordinate left and right sides by showing a preference for batting, kicking or throwing with one side or the other."

Kentucky educators used the British Columbia idea of widely held expectations as the basis for detailed Learning Descriptions of expected student development in the primary years (Boysen 1994). Learning Descrip-

tions, which are arranged in a seven-column continuum under the headings *beginning*, *developing*, *competent,* and *expanding*, currently exist for writing, reading, mathematics, independent learning/citizenship, arts and humanities, motor development, science, and social studies (Kentucky Department of Education, June 1994 and May 1995). Students whose skills cluster in the *competent* sections of a continuum should be able to perform successfully in that area in fourth grade.

While continua are useful ways to organize information, they can be misleading. The Kentucky Department of Education warns that students should not be expected to progress from one level to the next in a steady, orderly fashion (Boysen 1994). Professor of Education Judith Newman pointed out in her review of the British Columbia Primary Program that all children may not pass through the same sequence of stages when learning (Ministry of Education 1992). Educators and parents must remember that such continua are only general guides.

Local Normative Comparisons

Anderson and Pavan say parents also deserve approximate information on their child's relative standing in their school and class. However, rather than reporting the child's overall "position on the class totem pole," they urge teachers to focus on specific contexts.

> Children invariably develop on a "broken front," with higher achievements in some subjects or activities and lower achievements in others.... For a teacher to mentally average it out and simply inform the parent in global terms that their child is above, at, or below average is not only a lazy approach but also a disservice.

For example, reporting that Joan is one of the best readers, is currently lagging behind in arithmetic, and is near the middle in physical coordination is more helpful to parents than reporting that Joan is an average student overall, and is less problematic than reporting that she is a better student than Mary but not as good as Mark.

Criterion-Referenced Progress

Reporting progress with reference to exit criteria—criteria that must be met before a student can pass from one educational level to the next—answers the question, "Is a student on course or not to get to the destination on time?" This kind of question deemphasizes inappropriate comparisons with the rates of progress of classmates. "Being 'slow' or 'behind' is no longer highlighted. Rather, we chart the, perhaps modest, gains over time and worry only whether the trend is a happy one," writes Wiggins (1994).

Communicating such progress to parents is only possible if outcomes and performance criteria have been established as described in chapter 2. Mary Nall reported that as British Columbia's education-reform process continues to evolve, public concern about educational standards has contributed to an increasing emphasis on comparing student progress to the provincial learning outcomes rather than to the performance of other students. However, the widely held expectations are still used as helpful frameworks and are the main focus of reporting at the primary level (Ministry of Education, September 1994).

Written Progress Reports

Written reports and conferences play complementary roles in communicating student progress to parents. Using multiple media to report should increase effectiveness, because adults, like children, have different learning styles, and they will grasp graphic, textual, and auditory information with varying degrees of ease and speed.

Descriptive comments and rating scales are the chief components of most multiage written progress reports, though teachers in some multiage programs are required to assign grades. British Columbia's structured written reports rely solely on text to convey the aspects of progress identified above. Westmoreland School's Primary Progress Report, designed to fit the needs of a single school, combines graphic and textual elements: continua, comments, and rating scales. The Kentucky Department of Education offers a variety of reporting format options to its teachers, while stipulating the content that must be reported.

British Columbia's Structure Written Reports

Early in the province's education-reform process, the Ministry of Education adopted anecdotal comments as the best means of communicating student progress at the primary level.

> Although checklists are useful to teachers as a way of organizing or analyzing information, they are not appropriate as reporting devices. Checklists tend to fragment evaluation and to focus it on . . . bits and pieces of the whole. They fail to indicate the relative importance of each item to the others and to the whole. (1990b)

The Ministry directed teachers to describe "what the child can do"; the child's interests, attitudes, and learning needs; how the teacher was supporting the child's learning; and how parents might do so (1990b). Ministry publications presented examples and suggestions (1990c) and reported the experiences and advice of piloting teachers (1990a). Workshops were held

across the province. However, not all the province's teachers mastered anecdotal reports easily. According to Nall, the emphasis placed on "what a child *can do*" led some teachers to think, mistakenly, that they should only report on the positives.

Parent dissatisfaction soon surfaced. Major complaints were that anecdotal reports weren't clear enough, didn't provide external reference points, and didn't report problems honestly (Ministry of Education, Winter 1994). In November 1993, anecdotal comments were renamed "structured written reports" and specific guidelines were issued for writing them at each grade level. In grades 4-7, structured written reports were introduced, and letter grades were mandated after having been optional for years (Art Charbonneau 1993).

The current *Guidelines for Student Reporting* specify that reports should describe "what the student is able to do, areas of learning that require further attention or development, ways the teacher is supporting the students' learning needs." Teachers are also asked to suggest ways parents can help. The document presents many excellent suggestions for writing clear, concise reports that place performance in the contexts of developmental expectations and the provincial learning outcomes. It includes plentiful examples and a handy list of concise replacements for vague, verbose, or jargon-filled phrases. Teachers are required to provide parents with three written reports per year and with two or more informal reports, which may consist of conferences, telephone conversations, or written communications (Ministry of Education, September 1994).

Westmoreland School: Continua and Rating Scales

Westmoreland Elementary School's new Primary Progress Report (undated b) consists of four eleven-by-seventeen-inch sheets of pressure-sensitive paper. Parents receive one copy for each of the three reporting periods, and one copy goes into the school's permanent records. Each rating scale has spaces for fall, winter, and spring reporting periods. Space is designated for written comments each term, and the same continua are marked each term. The spring copy will thus show the cumulative marks and comments for all three terms.

Developmental continua are provided for reading, writing, spelling, math facts, math problem-solving, and small-muscle physical development. Each continuum is composed of eight boxes containing performance descriptions, linked by an arrow that begins as a thin line in the box at the far left and gradually thickens as it approaches the right side of the page. Teachers indicate mastery of a performance by placing an x at any point along the arrow, and use a slash to indicate partial mastery.

Considering the amount of the human brain that is devoted to processing visual information, presenting information graphically is a sensible, effective strategy. Marks along the length of a continuum are more easily interpreted by readers than long lists of criteria accompanied by columns of numbers or abbreviations, just as a graph is more quickly comprehended than a table of figures. Continua are excellent means of conveying to parents the developmental nature of learning. Comparing the positions of marks made at the ends of different terms reveals both individual growth and progress toward expected primary achievements. "The only thing that's tricky about it is that in reality, children's learning isn't necessarily as linear as the arrow," commented Snyder. As a result, x marks indicating complete mastery occasionally appear to the right of slashes.

Several rating scales augment the continua, including *characteristics of successful citizens/learners* with nine items, *reading/language arts* with ten items, and *math* with twelve items. Each item is rated "needs development," "developing," or "strongly in place." The teachers piloted the format during the 1994-95 school year, Carol Olson said, along with a similar report used for kindergartners. They revised the form after surveying parents last year and plan to further refine it if parental feedback and their own experience indicate further changes are desirable.

Kentucky: Educators Have Many Format Options

The Kentucky Education Reform Act sets educational objectives at the state level and mandates many of the means to be employed in meeting those objectives. However, the Kentucky Department of Education encourages local choice and supports variation in implementation (Gaustad 1992a). The area of reporting is no exception to this policy.

Teachers involved in the Pilot Phase of the Kentucky Early Learning Profile asked for a progress report compatible with the assessment system. A report was developed and used by the many additional teachers who participated in the field study phase of piloting during the 1993-94 school year. Many of these teachers, believing the form was "too open ended," said that a more structured report would be helpful. Four new, more structured progress reports with different formats were developed (Kentucky Department of Education, June 1994c), and information on how to use them was included in the KELP Teacher's Handbook (Boysen 1994).

All four progress reports are designed so that information collected and recorded in the KELP can be easily inserted in the appropriate places. Each format includes continua for reading, writing, mathematics, and independent learning and citizenship, and it provides space for teacher, student, and parent comments regarding both past progress and goals to work toward

Kentucky's Primary Assessment System

Kentucky's primary assessment system demonstrates how the elements of authentic assessment can be combined. The Kentucky Department of Education requires primary teachers to assess student learning by means of observations, anecdotal notes, conversations with parents and students, collecting student work samples and reflections on their own learning, and recording student performances that demonstrate specified subject skills. The state hired Advanced Systems in Measurement and Evaluation to design a portfolio-like system to help teachers collect, organize, and review this assessment evidence.

The deceptively short Kentucky Early Learning Profile (Kentucky Department of Education June 1994a)—four pages to be reproduced as needed—is accompanied by a thick handbook packed with explanations, practical suggestions, and examples (Boysen 1994). Parts A and B, a form for recording teacher conversations with parents and students and a diary of observations, were adapted with permission from the Primary Language Record developed by London's Centre for Primary Education. Part C, used for recording the specified nine types of performances, and D, a Learning Descriptions summary, were created to match Kentucky's goals and outcomes.

Every effort has been made to be responsive to teachers' needs and to accommodate different preferences, said Advanced Systems' Merry Denny. Hundreds of Kentucky teachers participated in field testing the KELP and recommending modifications. The KELP teacher's handbook suggests optional, alternative ways of using the system's various components. Finally, the Kentucky Department of Education allows each district, school, and teacher either to choose to use the KELP, in whole or in part, or to design a different primary assessment system that fulfills the assessment requirements.

during the next term. The continua use the same format as the Learning Descriptions. In fact, teachers may simply use Part D of the KELP instead of transferring the information. Teachers or schools may select one of these forms, use a district-designed form, or design their own progress report. Parents or guardians are to receive reports three times a year, in conjunction with conferences (Boysen 1994).

Conferences with Parents and Children

Written reports, however well-prepared and accurate, can only communicate bare-bones information about a student's progress. It takes person-to-person communication to put meat on those bones. In fact, the Kentucky Department of Education describes a progress report as merely "a summary of what teachers plan to discuss at the conference" (Boysen 1994).

Conferences provide opportunities to build rapport with parents, educate them about instructional and assessment practices, prevent or clear up misconceptions, and obtain information about how the student's family supports his or her learning. They also enable teachers to communicate sensitive information or tentative judgments best not put in writing.

Children should not be excluded from a process that concerns them so profoundly. In some programs, teachers conference with students before the parent conference. In others, students participate in three-way conferences

with parents. The Bronx New School invites all family members to attend conferences. This helps link home events and school learning, builds a sense of community, and yields additional information about students. "Sometimes even a sibling can provide insights into the learning style or behavior of a student," Falk reports.

A Model Conference Cycle

"The parent-teacher conference is the approach most universally advocated in the current literature of reporting and is probably the most fruitful and effective single means available," wrote Goodlad and Anderson in their pioneering work *The Nongraded Elementary School*. Three decades later, Anderson and Pavan reiterate the statement. They suggest the following five-stage sequence as the "ideal" conferencing cycle.

Stage 1: The "home base" teacher shares information and perceptions with the other professionals who work with the child, and the child is given opportunities to review and reflect on his or her progress.

Stage 2: The teacher and child share their perceptions of the student's progress and agree on "things to celebrate and things to do next." The student may be asked to prepare a portfolio of work for the teacher and parents to view.

Stage 3: The teacher, parents, and possibly the child meet and share information. Past accomplishments are acknowledged, problems are identified, and goals for the future are set.

Stage 4: Parents and children discuss these issues at home.

Stage 5: Children and parents give the teacher feedback regarding the conference and the issues covered.

Anderson and Pavan acknowledge that, given the social conditions of the 1990s, this ideal cycle may not always be possible. If parents are unable or unwilling to come to the school, the authors suggest teachers go through stages 1 and 2, communicate the information to parents via phone or in writing, and encourage parents and children to carry out stages 4 and 5, providing feedback to the teacher after discussing the issues at home.

Conferencing Suggestions

NCRESST (undated a) recommends including the following basic elements in every conference. Parents should be asked for their perceptions of students' strengths and weaknesses, and students should be asked to assess their own strengths and weaknesses. Teachers should discuss "*multiple* indicators of student performance" in academic areas, as well as students' progress toward mastering social skills such as working cooperatively and assuming leadership roles. Work samples that "demonstrate a wide range of

problem-solving skills" and reveal areas where improvement is needed should be shown and discussed.

Ministry of Education publications give numerous tips for preparing for, conducting, and following up after conferences. Asking parents to fill out surveys ahead of time, listing topics they would like to discuss, can help teachers select key areas to focus on in the conference. Everything can't be covered in one meeting, and parents may feel overwhelmed if too much information is presented at once. The teacher should begin with a positive comment, avoid using educational jargon that can confuse or intimidate parents, and illustrate points with specific examples of student behavior. Teachers should ask parents open-ended questions, take notes, listen carefully, and rephrase parents' statements to ensure they have correctly understood them. After the conference, parents should be sent a note thanking them for attending and asking them for feedback (September 1994, 1990c).

If conferences are three-way, teachers can discuss the conference process with the class beforehand and have students practice their presentation by role-playing with each other. The student can introduce parents and teacher at the beginning of the conference, show parents the classroom and his or her portfolio, present a self-report, and comment on how it compares with the teacher's report. After all the conferences are finished, the teacher can discuss the process with the entire class, record suggestions for the next conference cycle, and ask each child to fill out a conference evaluation (Ministry of Education 1990c).

Conferences Replace Report Cards at Morse Street School

Teachers at Morse Street School in Freeport, Maine, have discovered a way to guarantee congruence among curriculum, assessment, and reporting: They use the same document as the basis for all three activities. The two parent conferences the district requires each year are spent discussing the wealth of information recorded in the nearly fifteen-page-long assessment record, which remains at the school. The only written report parents receive is a single anecdotal report at the end of the year.

The assessment instrument contains detailed, comprehensive developmental checklists for writing, literacy, math, and physical-motor development, plus two sets of continua. Social-emotional development is assessed with ten continua for specific capabilities such as cooperation, self-control, and problem-solving, and the area of cognitive reasoning with five continua (Morse Street School 1995). "I explain to parents, it's as if we're showing them the teacher's working plan book for their child," said Principal Cheryl White. "Everything the teacher keeps track of is on there. That's why it doesn't go home."

Teachers organize conferences in slightly different ways, said White. Some teachers put the assessment documents in the children's portfolios, leave them in the waiting area outside the classroom, and ask parents to arrive fifteen minutes early to review their child's portfolio, looking for the marks whose color corresponds to their child's year in the program. Marks made on the assessment document are color-coordinated with the portfolios: blue for kindergarten, green for first grade, and red for second grade. Some teachers go over the documents with each child before the parent conference, while in other cases children are included in the parent conference.

Teachers spend from half an hour to forty-five minutes going through the document with each set of parents, showing them student work samples that illustrate each point. White explained, "A teacher might say, 'I marked here that your child mastered inventive spelling, and right here's a paper that shows what I mean. And please take this work home with you'." Taking these work samples home from the conference seems to satisfy most parents' desire for documentation.

The fact that parents only see the document with the teacher at hand, ready to answer questions, prevents misunderstandings from taking root. For example, when parents of kindergartners first see the many pages of checklists and continua, their reaction is often, "My gosh, how can my child do all this in one year?" The teacher can quickly reassure them, explaining that the document covers the entire K-2 spectrum of development.

Because the new reporting process is so much more time-consuming than the previous one, even without the work of preparing progress reports each term, the school reduced the number of reporting periods from four to three per year. This is a common schedule for multiage programs with developmentally appropriate reporting schemes. Conferences occur in November and March, and the anecdotal report is sent home in June.

Conclusion

Changing reporting practices is a long-term process in which psychological barriers often loom larger than technical ones. Many parents will find it very difficult to let go of traditional grades and fully accept developmentally appropriate reporting methods. Teachers and administrators must be extremely sensitive to these strongly held, often irrational attitudes as they introduce reporting changes.

Educators who change reporting practices too rapidly may be forced to make a hasty retreat. Wiggins (1994), Olson (1995), and Seeley cite instances where parents demanded the reinstatement of grades or the addition of some type of norm-referenced information to progress reports. In extremely traditional communities, retaining grades as an adjunct to develop-

mentally appropriate reporting methods until parents become familiar with the new methods might be a wiser course of action than eliminating grades abruptly and suffering a violent backlash.

Wiggins holds that letter grades need not be harmful, especially if they are just one element of a report card that is rich in scores for different subtypes of achievement, uses rubrics with clearly defined, agreed-upon criteria rather than subjective judgment, or adds narrative descriptions of performance accompanied by student work samples. Alfie Kohn, a fierce opponent of letter grades, suggests ways to diminish their negative effects while they still persist, such as assigning grades only at the end of the term and never grading on a curve.

Attitudes about grading practices may change slowly and with difficultly, but they can change, as the experience of Kentucky educators demonstrates. Denny reports that in schools where teachers haven't fully accepted the reforms themselves, parents remain suspicious of the new reporting methods and believe grades have more validity. "Their comfort level is with grades, and they want grades." However, where teachers understand authentic assessment and are successfully using KELP with multiage instructional methods, "parents come to the conferences and love everything they get. And then, at the very end, they kind of lean over and *whisper*, "But if you were going to give a letter grade, what letter grade would you give my child?"

"At least they whisper now!" said Denny with a laugh. "They are learning to do without grades, although they would still like them. Maybe in a few years they won't even whisper, they just won't ask."

Chapter 4

Implications for School Boards and Administrators

Administrators and school board members have two main roles to play in relation to authentic assessment: to support teachers, and to promote understanding and acceptance of developmentally appropriate assessment and reporting practices in the community. The first responsibility falls mainly to principals and other building-level administrators. The second responsibility is shared by school board members, district-level administrators, principals, and teachers. Both tasks are essential for authentic assessment to succeed, whether change is mandated from above or initiated by local teachers and administrators dissatisfied with conventional assessment methods.

Helping Teachers Implement Authentic Assessment

It is a tremendous challenge to design and implement authentic-assessment methods, even for educators who are eager to try such methods. Grant Wiggins comments, "Teachers now face the same situation as students who are asked to do a nonroutine task: how do they design new forms of assessment, the likes of which they've never seen? They know what they don't like about conventional testing... but they don't know what this new vision looks like" (Ron Brandt 1992).

Teachers need time, staff development, and long-term technical and psychological support as they learn the skills they need to make this vision a reality. These types of support are likely to require extra financial resources. Principals ideally should have excellent interpersonal, management, and fund-raising skills as well as understanding of, and commitment to, authentic assessment.

Assessment Must Be Part of Comprehensive Change

Administrators who want to promote authentic assessment must support comprehensive changes in curriculum and instruction, though multiage grouping itself is not essential. According to Merry Denny, Kentucky teachers who had not made organizational and instructional changes had great difficulty implementing the KELP'S authentic-assessment methods. "In almost every single case, the teachers who could not use the KELP had not made changes," said Denny. "And those people who had made changes had very little trouble using the KELP. It fit."

Conversely, authentic assessment facilitates the implementation of developmentally appropriate instructional practices. Kentucky teachers tended to fall into three groups, Denny explained: those convinced of the value of the new methods, those adamantly opposed to change, and those in the middle—often the largest group in any major change process. It is difficult to change the attitudes of the antichange faction, said Denny. "But for the middle group, who didn't completely understand the reforms but were trying to change, we found that the KELP was a very strong mover. It helped them to focus on appropriate instructional strategies and techniques."

An as yet unpublished study conducted by University of Louisville researchers concluded that using the KELP or KELP-like systems helped Kentucky teachers better understand children's development and the link between authentic assessment and developmentally appropriate instructional practices, said Ric Hovda, director of the Center for the Collaborative Advancement of the Teaching Profession at the University of Louisville School of Education. A study in Vermont, which mandates portfolio assessment at the fourth- and eighth-grade levels, also found a relationship between changes in assessment and instruction: Teachers and administrators reported that portfolio use resulted in changes in instructional strategies and curriculum content (Herman and Winters).

Staff Development and Ongoing Support

According to Shepard (1995), supporters of assessment reform often underestimate the extent and depth of staff development that will be required. She concludes that staff development should provide teachers with appropriate materials to explore and adapt to individual needs; time to reflect and plan; opportunities to discuss ideas and share experiences with colleagues; and ongoing support from experts as they learn both practical techniques and the conceptual bases for them.

One source of appropriate materials is a database compiled by NCRESST (undated a) that includes listings of over 250 alternative assessments developed across the nation. Other sources include professional jour-

nals, conferences, state assessments, and other schools and districts. Herman and her colleagues encourage educators to build on the work others have done, borrowing and combining elements of existing materials to create assessments that suit their particular needs and purposes.

Blaine R. Worthen (1993b) advises educators to seek out local sources of expertise: "There is something deeply comforting about having someone nearby who 'has been there' and can demonstrate new methods to neophytes or critique their embryonic efforts." The Kentucky Department of Education encourages such teacher-to-teacher support by bringing together teachers with varying levels of assessment expertise at regional meetings. As Denny explained,

> Those who were having problems would say, "I can't *do* this, this isn't *possible*." Then teachers who were farther along in the process would say, "You know, last year I thought the exact same thing. And look, I did it this way. Why don't you go back and try this."

Support should be provided on an ongoing basis, not just in one-time workshops at the start of implementation. In Shepard's assessment project (1995), specialists held weekly workshops with the participating teachers. Shepard found that expert advice often does not make sense to teachers until they have had opportunities to experiment with techniques in the classroom.

Teachers need to understand the philosophical and conceptual bases of authentic-assessment as well as the methods themselves. Without indepth understanding, teachers may unintentionally distort or oversimplify assessments as they put them into practice. They will also have difficulty explaining to parents the reasons for using authentic-assessments (Shepard 1995).

White believes that supporting teachers as they explore new methods and set new assessment goals is much more effective than pushing them to achieve goals established by others.

> I think too many administrators hop on a bandwagon and say, "This is what we're going to do." But I think that does a disservice to everybody involved. I'm firmly convinced that most public schools are full of wonderful teacher talent. As an administrator, you need to give them room to develop that talent. And if you don't go through the process of involving teachers every single step of the way, it gets put in a drawer and nobody does it. If they don't own it, forget it.

A recent performance-assessment study of fourteen schools in thirteen states supports White's opinion. In general, the study found classroom-level changes to be positive, but minimal; changes were most notable in schools where teachers "had been involved with the new assessment systems from the start" (Debra Viadero 1995).

Another way to support teachers is to remove the pressure of standardized testing, at least while authentic assessments are in the initial stages of implementation. Shepard's researchers encouraged teams of teachers to

participate in their assessment project by obtaining two-year waivers from standardized testing by the state. This involved obtaining approval from "district officials, the teachers union, and each school's parent accountability committee" (Shepard 1995).

Time and Money

"Changing assessment takes more time than you expect," commented Spizzuoco, one of the five members of Morse Street School's assessment committee. Her conclusion is corroborated by many authentic-assessment studies. A panel of testing experts concluded that Kentucky "tried to do 'too much, too fast,' and therefore has spent a lot of time trying to patch up problems" in its new assessment system (Lonnie Harp 1995). Researchers in Shepard's study (1995) had to slow the pace to give teachers time to absorb and implement all the information they were receiving. Educators must be aware that practices described in a few brief paragraphs in this Bulletin may take months or years to implement successfully. Creating consensus on outcomes and performance criteria is also time-consuming.

Even after teachers are comfortable with authentic-assessment methods, using them continues to require more time than using conventional methods. A Vermont study of portfolio use found that selecting portfolio tasks, preparing related lessons, and evaluating portfolio contents took teachers seventeen hours per month, and 60 percent of the teachers surveyed said they often didn't have sufficient time to prepare lessons (Herman and Winters).

In a 1993 study of forty-six Kentucky primary schools, researchers found that few teachers were actually conducting the required parent conferences. "They explained that there was no time built into the school schedule for conferences, and they were not compensated for the extra time... so they were not conducting conferences on a regular basis" (Institute on Education Reform 1994). Providing teachers with the time they need may be costly, but it is essential if authentic assessment is to be fully implemented.

Examining the history of Morse Street School's assessment project shows the interwoven roles of time and money. White subsidized the cost of staff time by obtaining a $5,000 grant from the UNUM Insurance company through the Southern Maine Partnership, an association of schools coordinated by the University of Southern Maine. She used the money to hire substitutes and to pay teachers hourly stipends for attending meetings during the summer and after school. When the project took longer than expected, a second grant was obtained, and work on the project continued without grant support for two additional years. The project took four years to complete instead of the one year originally anticipated (Morse Street School).

Designing the new assessment system required substantial blocks of quality time, said Spizzuoco: "It's just not something that's easy to do after school and in between." The grant enabled the assessment committee to get a head start on the project during the summer of 1991. During the school year, the two normally scheduled staff development days were devoted to assessment, and White created additional blocks of time by hiring substitutes for special assessment days. Members of the assessment committee were released for the entire day, and five substitutes rotated from room to room, releasing the remaining teachers five at a time to meet with the committee for one to two hours. "That was the best way to do it and not do it at 3:00, when people just didn't feel like talking," Spizzuoco explained.

This schedule did have some drawbacks, however, Spizzuoco said. "Sometimes the first group would say something, and then the third group would say something completely different, and we knew we needed to bring it up to everyone." She advises large schools planning assessment changes to arrange for their faculties to meet both in small groups and as a whole.

White made special efforts to ease the load on teachers and boost morale. On assessment days she took the five rotating substitutes to her office and trained them to present a lesson. That way, teachers didn't have to spend extra time preparing a lesson for the substitute to teach. "I wanted to make it as easy as possible for them," she explained. "And I always had food and chocolate at the meetings."

Educating and Involving the Community

As educators in Cranston, Rhode Island, discovered, years of careful work can be suddenly jeopardized by negative community reactions to assessment and reporting changes (Olson 1995). Educating the community must be a high priority in any plan to implement authentic assessment. However, this is too large a topic to be discussed in depth here.

Many administrators and school board members will need to begin this process by making themselves more assessment literate. They cannot be effective spokespersons for authentic assessment unless they understand the shortcomings of conventional assessments and the potential of authentic assessment. Their goal should be to become open-minded but critical consumers of all methods, remembering that authentic assessment is still in its infancy, especially its large-scale, high-stakes forms.

Working with the Media

In working to increase assessment literacy in the community, policymakers will have to fight the common American tendency to oversimplify.

"American consumers of information want assessment data reduced to very small, ostensibly easy-to-understand chunks that lend themselves to the reporting formats of newspapers and television," comments Lescher.

Changing this public mindset will be a long, slow process. Administrators and policy-makers can begin by reporting more complex information, such as explaining how standardized test scores should be interpreted as well as reporting the scores, pointing out the tests' inherent limitations and margins of error rather than treating them with uncritical awe, and reporting other types of assessment evidence, too. They should firmly but patiently insist that local media report more complex and varied assessment information, and report it in context.

Involving Parents and Members of the Community

Parents and members of the community will learn much more about authentic assessment if they are actively involved in the process. As was mentioned in chapter 2, members of the community can be asked to participate in setting learning goals. Herman and others suggest involving business representatives and other citizens in "generating real-world, authentic" performance-assessment tasks.

Clyde Miller Elementary offers authentic-assessment training at intervals throughout the year to parents and interested community members. Once trained, the adults join a pool of raters periodically asked to assist teachers in evaluating student projects. "The crew at Fire Station #5; employees from our business partner, SuperValu Grocery Warehouse; and staff at Aurora Community College are all valued raters at our school," report Kenney and Perry.

Local employers could be asked to support assessment by providing employees with paid time off for parent conferences during school hours. Anderson and Pavan suggest asking local service organizations such as Rotary, Kiwanis, and Chambers of Commerce to help promote this idea.

Conclusion

Authentic assessments provide teachers of multiage classes with compatible, effective methods for assessing, evaluating, and reporting student progress. In fact, using authentic assessment should make multiage teaching easier. Authentic assessments can also be used in age-graded classes that use developmentally appropriate instructional practices. Both authentic assessment and developmentally appropriate instructional practices are becoming increasingly common in classrooms across the continent, despite occasional difficulties and setbacks.

Problems of validity, reliability, generalizability, and expense still remain to be resolved with respect to large-scale, high-stakes performance assessments. These problems are much less relevant to classroom-level authentic assessments. However, even methods with great potential can be sabotaged by hasty or poorly planned implementation. Particular care must be taken to inform and involve parents during the process of changing assessment and reporting approaches.

"No school should squander its opportunity by plunging precipitously into an alternative assessment effort before it is ready to do it well," cautions Worthen (1993b). Educators should proceed carefully but optimistically, aware of both the strengths and weaknesses of authentic assessment and aware that implementation will take considerable time. The evidence to date suggests that authentic assessment is well worth the effort of doing it well.

Appendix

Authentic Assessment in Oregon

The national trend toward performance-based assessment is very evident in Oregon. In the 1990-91 school year, Oregon implemented statewide criterion-referenced assessments linked to the state's Essential Learning Skills and Common Curricular Goals (Oregon Department of Education 1992). "Open-ended" mathematics assessments, in which problems "have more than one possible solution, require multiple steps to solve, and require students to explain the steps they took to solve the problems," were given to fourth- and eighth-grade students for the first time during the 1994-95 school year (Oregon Department of Education undated).

Oregon is also participating in the national New Standards Project, a voluntary association of states working to develop "a national examination system that reflects international standards of performance." The proposed examination system would include a Performance Examination component and a Cumulative Accomplishments component (Oregon Department of Education 1992).

Revisions to the Oregon Educational Act for the 21st Century

Revisions made to the Oregon Educational Act for the 21st Century by the 1995 legislature reaffirmed the element of performance assessment. The revised act instructs the State Department of Education to update the Common Curriculum Goals, to develop "criterion-referenced assessments including performance-based, content-based and other assessment mechanisms to test knowledge and skills," and to establish criteria for the revised outcomes for the Certificates of Initial Mastery and Advanced Mastery and for benchmarks at grades 3, 5, 8 and 10 (Oregon Legislative Assembly 1995).

Department of Education staff are in the process of setting these content and performance standards, with input from educators, and developing statewide assessments in mathematics, science, English, history, geography, economics, and civics. Existing statewide assessments will ultimately be integrated with the performance assessments schools will use to assess student progress toward the certificates (Oregon Department of

Education, undated). "We want to make the large-scale assessments as performance-oriented as we possibly can," said Barbara Wolfe, coordinator of assessment for the department. Districts will be required to develop their own performance standards and assessments for the arts and second languages (Oregon Department of Education 1995).

Wolfe commented on the state's changing assessment role. "In the past, our purpose was to provide schools with information for program improvement. Now our purpose is changing to focus more on individual students. The statewide assessment will be an important factor in determining whether a student is making progress toward the Certificate of Initial Mastery." Wolfe acknowledges that it will be quite a challenge to develop assessments that are equally appropriate for students throughout the state. Great variation exists among school districts in different Oregon communities because of Oregon's strong tradition of local management. "Of course our intent is not to compare schools, but inevitably the data are used for that by the media and others. So we have to be cognizant of that, even though that's not our primary purpose," she added.

At the primary level, Wolfe considers it appropriate for state-administered tests to check student progress in literacy and mathematics. However, she thinks other content might be better assessed by classroom teachers, provided that they receive good staff development. "I would like to see the state do more to assist schools in using appropriate performance-based, observational assessment tools," she said.

New State Writing Assessments Match Changes in Instruction

In 1989, the Oregon Department of Education received funding to expand the statewide direct writing assessment program. After a successful 1990 pilot, all Oregon third-, fifth-, eighth-, and eleventh-graders participated in the assessment in 1991 (Wolfe and others). The performance-based writing assessments (described in chapter 1) are revealing improvements in student writing proficiency that conventional assessments might have overlooked, as well as areas where improvement is still needed. According to Wolfe, changes in the teaching of writing in the classroom underlie the changes in test scores.

Scores for 1994 revealed significant growth in the writing performance of third-grade students, said Wolfe. In the area of ideas and content—"the ability to identify a topic, stick with the topic, and develop it with adequate supporting detail"—the percentage of third-graders scoring in the top two levels of proficiency rose from 21.8 percent in 1991 to 31 percent in 1994, a jump of nearly ten points, while the percentage scoring in the lowest two levels dropped more than 8 percent. An increase of similar size occurred in the area of voice, and smaller, but still welcome, increases occurred in most other areas.

Improvement among eighth-graders was even more impressive. Scores rose significantly in every area of writing

performance. The smallest increase in the percentage of students scoring at the upper two levels—from 31.1 to 39.9 percent—occurred in the area of *conventions*, which includes spelling, grammar, capitalization, punctuation, and usage. Increases of more than 20 percentage points occurred in several areas. In *ideas and content* the percentage of high-scoring eighth-graders jumped from 25 percent in 1991 to a whopping 46.5 percent in 1994 (Oregon Department of Education, undated).

Wolfe believes these improvements are related to the increased use of process-writing approaches and developmentally appropriate practices in Oregon classrooms. She explained, "Over the last ten to twelve years in Oregon we've seen a tremendous increase in the number of teachers who are using writing-as-a-process approaches and having young children write instead of fill in blanks." Allowing students to choose topics they are interested in to write about, a practice more common in developmentally appropriate classrooms than in traditional ones, also contributes to improved student writing, she said.

The statewide assessment revealed one notable problem area among third-graders: the percentage of students with high scores in *conventions* dropped more than ten points. Wolfe thinks several factors may have contributed to the drop in this area.

First, many third-grade students are attempting more than their predecessors did, said Wolfe. "Just by measuring with a ruler we can prove they're writing more than they used to write," she commented, "and the more you write, the more opportunity for error." Students are also attempting to use more sophisticated language and to write more complex and difficult kinds of text, such as dialogue. Many studies have shown that errors in language usage increase as learners attempt more difficult tasks, said Wolfe.

Another factor may be that teachers are devoting less time and attention to the conventions of language than they did in the past. "That may be partly because they're trying to teach aspects of writing that are more difficult to teach. Teaching how to develop an idea is harder than teaching what words to capitalize." As teachers focus on mastering these more complex teaching tasks, they may temporarily underemphasize conventions. It is also possible that some teachers have misinterpreted some aspects of process-writing theory, said Wolfe, and mistakenly think "that being developmentally appropriate means you don't ever give direct instruction, or don't correct errors. I have encountered some teachers who mistakenly came to believe that."

Both increases and decreases provide useful information for administrators, said Wolfe.

> If I were a building administrator and my building's scores paralleled these numbers, that would tell me what conversation to have with the primary teachers: how can we keep our students writing well, organizing well, using their wonderful, strong, individual voices to tell their stories, and also help them to correctly use the conventions as they attempt these more difficult, complex types of text.

Bibliography

Anderson, Robert H., and Barbara Nelson Pavan. *Nongradedness: Helping It To Happen.* Lancaster, Pennsylvania: Technomic Publishing Co. Inc., 1993. 266 pages. ED 355 005.

Batzle, Janine. *Portfolio Assessment and Evaluation: Developing and Using Portfolios in the Classroom.* Cypress, California: Creative Teaching Press Inc., 1992. 125 pages.

Boysen, Thomas C. *Kentucky Early Learning Profile: A Part of the Kentucky Primary Assessment System, Teacher's Handbook.* Frankfort, Kentucky: Kentucky Department of Education, June 1994. 192 pages.

_____. *State Regulations and Recommended Best Practices for Kentucky's Primary Program.* Frankfort, Kentucky: Kentucky Department of Education. January 1993. 25 pages.

Brandt, Ron. "On Misuse of Testing: A Conversation with George Madaus." *Educational Leadership* 46, 7 (April 1989): 26-29. EJ 387 140.

_____. "On Performance Assessment: A Conversation with Grant Wiggins." *Educational Leadership* 49, 8 (May 1992): 35-37. EJ 444 310.

Bridge, Connie A.; Beverly A. Reitsma; and Peter N. Winograd. *Primary Thoughts: Implementing Kentucky's Primary Program.* Lexington, Kentucky: Kentucky Department of Education, 1993. 264 pages. EJ 370 678.

Brown, Rexford. "Testing and Thoughtfulness." *Educational Leadership* 46, 7 (April 1989): 31-33. EJ 387 141.

Caine, Renate Nummela, and Geoffrey Caine. *Making Connections: Teaching and the Human Brain.* Menlo Park, California: Innovative Learning Publications, Addison-Wesley Publishing Company, 1994. 214 pages.

Calkins, Trevor. "Off the Track: Children Thrive in Ungraded Primary Schools." *School Administrator* 49, 5 (May 1992): 8-13. EJ 444 301.

Charbonneau, Art. *Improving the Quality of Education in British Columbia: Changes to British Columbia's Education Policy.* Victoria, British Columbia: Ministry of Education, November 1993. 9 pages.

Chase, Pennelle, and Jane Doan. *Full Circle: A New Look at Multiage Education.*

Portsmouth, New Hampshire: Heinemann, 1994. 184 pages. ED 371 864.

Clay, Marie. *An Observation Survey of Early Literacy Achievement.* Portsmouth, New Hampshire: Heinemann, 1993. 93 pages. ED 358 427.

Darling-Hammond, Linda. "Setting Standards for Students: The Case for Authentic Assessment." *NASSP Bulletin* 77, 556 (November 1993): 18-26. EJ 472 577.

De Fina, Allan. *Portfolio Assessment: Getting Started. Teaching Strategies.* Jefferson City, Missouri: Scholastic, Inc. 1992. 88 pages. ED 371 339.

Diegmueller, Karen. "Bucking a Trend, Maryland Assessments Emerge as Model School-Reform Tool." *Education Week* XIV, 41 (August 2, 1995a): 12-13.

_____. "Lawmakers Putting the Brakes on Alternative Assessments." *Education Week* XIV, 41 (August 2, 1995b): 12-13.

_____. "Push Ahead with Standards, Alternative Assessments, Panel Urges." *Education Week* XIV, 41 (August 2, 1995c): 6-7.

Fairplay Elementary School. *History of Fairplay.* Corvallis, Oregon: Author, February 1994. 9 pages, unpaginated.

_____. *Oregon Business Education Fellowship Final Evaluation.* Corvallis School District 509J. Corvallis, Oregon: Author, undated. 5 pages.

Falk, Beverly. *The Bronx New School: Weaving Assessment into the Fabric of Teaching and Learning.* New York: National Center for Restructuring Education, Schools, and Teaching, September 1994. 51 pages. ED 378 217.

Fredericks, Anthony D, and Timothy V. Rasinski. "Involving Parents in the Assessment Process." *The Reading Teacher* 44, 4 (December 1990): 346-49. EJ 418 099.

Gardner, Howard. *The Unschooled Mind: How Children Think and How Schools Should Teach.* New York: Basic Books, 1991. 303 pages. ED 368 451.

Gaustad, Joan. *Making the Transition to Nongraded Primary Education.* Oregon School Study Council Bulletin Series. Eugene, Oregon: Oregon School Study Council, April 1992. 41 pages. ED 342 282.

_____. *Nongraded Education: Mixed-age, Integrated, and Developmentally Appropriate Education for Primary Children.* Oregon School Study Council Bulletin Series. Eugene, Oregon: Oregon School Study Council, March 1992. 38 pages. ED 343 227.

_____. *Nongraded Education: Overcoming Obstacles to Implementing the Multiage Classroom.* Oregon School Study Council Bulletin Series. Eugene, Oregon: Oregon School Study Council, November and December 1994. 84 pages. ED 379 744.

Goodlad, John I., and Robert A. Anderson. *The Nongraded Elementary School..* Revised Edition. New York: Teachers College Press, Columbia University, 1987. 248 pages. ED 279 084.

Grant, Jim, and Bob Johnson. *A Common Sense Guide to Multiage Practices.* Petersborough, New Hampshire: Crystal Springs Books, 1994. 124 pages. ED 373 915.

Guskey, Thomas R. "A Look Back at Grading Practices." *Educational Leadership* 52, 2 (October 1994a): 18.

_____. "Making the Grade: What Benefits Students?" *Educational Leadership* 52, 2 (October 1994b): 14-20.

Haladyna, Thomas M.; Susan Bobbit Nolen; and Nancy S. Haas. "Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution." *Educational Researcher* 20, 5 (June-July 1991): 2-7. EJ 432 068.

Haney, Walter, and George Madaus. "Searching for Alternatives to Standardized Tests: Whys, Whats, and Whithers." *Phi Delta Kappan* 70, 9 (May 1989): 683-87. EJ 388 720.

Harp, Lonnie. "Kentucky Student Assessments Called 'Seriously Flawed'." *Education Week* XIV, 40 (July 12, 1995): 12-13.

Herman, Joan L. "What Research Tells Us About Good Assessment." *Educational Leadership* 49, 8 (May 1992): 74-78. EJ 444 324.

Herman, Joan L.; Pamela R. Aschbacher; and Lynn Winters. *A Practical Guide to Alternative Assessment.* Alexandria, Virginia: Association for Supervision and Curriculum Development, 1992. 124 pages. ED 352 389.

Herman, Joan L., and Lynn Winters. "Portfolio Research: A Slim Collection." *Educational Leadership* 52, 2 (October 1994): 48-55.

Institute on Education Reform, University of Kentucky. *Kentucky's Primary Program: A Progress Report.* Lexington, Kentucky: Authors, February 1994. 56 pages. ED 369 517.

Katz, Lilian G. *A Developmental Approach to the Education of Young Children: Principles of Assessment.* Urbana, Illinois: ERIC Clearinghouse on Elementary and Early Childhood Education, June 25, 1991. 2 pages.

Kenney, Evelyn, and Suzanne Perry. "Talking with Parents About Performance-Based Report Cards." *Educational Leadership* 52, 2 (October 1994): 24-27.

Kentucky Department of Education. *Kentucky Early Learning Profile.* Frankfort, Kentucky: Author, June 1994a. 5 pages.

_____. *Kentucky Early Learning Profile Learning Descriptions: The Arts and Humanities, Motor Development, Science, Social Studies.* Working Document. Frankfort, Kentucky: Author, May 1995. 43 pages.

_____. *Kentucky Early Learning Profile Learning Descriptions: Writing, Reading, Mathematics, and Independent Learning/Citizenship.* Working Document. Frankfort, Kentucky: Author, June 1994b. unpaginated.

_____. *Progress Report To Be Used in Conjunction with the Kentucky Early Learning Profile.* Frankfort, Kentucky: Author, June 1994c. 4 pages.

Kohn, Alfie. "Grading: The Issue Is Not How but Why." *Educational Leadership* 52, 2 (October 1994): 38-41.

Lescher, Marianne Lucas. *Portfolios: Assessing Learning in the Primary Grades.* Washington, D.C.: National Education Association Professional Library, 1995. 40 pages.

Madaus, George F., and Thomas Kellaghan. "The British Experience with 'Authentic' Testing." *Phi Delta Kappan* 74, 6 (February 1993): 458-59, 462-63, 466-69. EJ 457 279.

Maeroff, Gene I. "Assessing Alternative Assessment." *Phi Delta Kappan* 73, 4 (December 1991): 272-281.

McCann, Julie. *Promoting Teacher Involvement in Assessment, Change, and Improvement.* Corvallis, Oregon: Fairplay Elementary School, Corvallis School District 509J, undated. 52 pages.

Miller, Laura. "Early-Years Educators Out Front in Ditching Traditional Tests." *Education Week* XIV, 32 (May 3, 1995): 8-9.

Ministry of Education, Province of British Columbia. "British Columbia Provincial Curriculum Speaker's Notes." Victoria, British Columbia: Field Services Team, September 12, 1995. 21 pages.

_____. *Charting Change: Review of the Primary Program in British Columbia, Interim Report II.* Victoria, British Columbia: Author, Winter 1994. 73 pages.

_____. *Guidelines for Student Reporting for the Kindergarten to Grade 12 Education Plan.* Victoria, British Columbia: Author, September 1994. 43 pages.

_____. *Mathematics K to 7: Integrated Resource Package 1995.* Victoria, British Columbia: Author, 1995a. 263 pages.

_____. *Our Primary Program: Taking the Pulse.* Victoria, British Columbia: Author, 1990a. 170 pages.

_____. *Personal Planning K to 7: Integrated Resource Package 1995.* Victoria, British Columbia: Author, 1995b. 203 pages.

_____. *Primary Program Foundation Document.* Victoria, British Columbia: Author, 1990b. 362 pages.

_____. *Primary Program Resource Document.* Victoria, British Columbia: Author, 1990c. 392 pages.

_____. *Science K to 7: Integrated Resource Package 1995.* Victoria, British Columbia: Author, 1995c. 263 pages.

_____. *Supporting Learning: Understanding and Assessing the Progress of Children in the Primary Program. A Resource for Parents and Teachers.* Victoria, British Columbia: Author, 1991. 62 pages.

Ministry of Education and Ministry Responsible for Multiculturalism and Human Rights, Province of British Columbia. *Building Firm Foundations: Review of the Primary Program in British Columbia, Interim Report I.* Victoria, British Columbia: Author, Fall 1992. 32 pages.

Morse Street School. *Assessment Handbook for the Early Childhood Center.* Working draft. Freeport, Maine: Author, 1995. 49 pages.

National Association for the Education of Young Children. *Developmentally Appropriate Practice in Early Childhood Programs Serving Children From*

Birth Through Age 8. Edited by Sue Bredekamp. Washington, D.C.: Author, 1987. 91 pages.

_____. "NAEYC Position Statement on Standardized Testing of Young Children 3 Through 8 Years of Age." *Young Children* 42, 3 (March 1988): 42-47.

National Association for the Education of Young Children and National Association of Early Childhood Specialists in State Departments of Education. "Guidelines for Appropriate Curriculum Content and Assessment in Programs Serving Children Ages 3 Through 8." *Young Children* 46, 3 (March 1991): 21-38.

National Center for Education Statistics. *Interviewing Children About their Literacy Experiences: Data from NAEP's Integrated Reading Performance Record (IRPR) at Grade 4.* Washington, D.C.: Office of Educational Research and Improvement, January 1995a. 92 pages.

_____. *Windows into the Classroom: NAEP's 1992 Writing Portfolio Study.* Washington, D. C.: Office of Educational Research and Improvement, January 1995b. 272 pages.

National Center for Research on Evaluation, Standards and Student Testing. *Guidebook for the Video Program "Assessing the Whole Child."* Los Angeles, California: Author, undated a. 21 pages.

_____. *Guidebook for the Video Program "Portfolio Assessment and High Technology."* Los Angeles, California: Author, undated b. 24 pages.

Neill, D. Monty, and Noe J. Medina. "Standardized Testing: Harmful to Educational Health." *Phi Delta Kappan* 70, 9 (May 1989): 688-97. EJ 388 721.

Olson, Lynn. "On Assignment: Cards on the Table." *Education Week* XIV, 38 (June 14, 1995): 23-28.

Oregon Department of Education. *Academic Achievement.* Salem, Oregon: Author, undated. 7 pages.

_____. *1995 Update: Oregon Educational Act for the 21st Century.* Salem, Oregon: Author, 1995. 65 pages

_____. *Oregon Report Card: An Annual Report on Public Education to the Citizens of Oregon.* Salem, Oregon: Author, Fall 1992. 74 pages.

Oregon Legislative Assembly. *Enrolled House Bill 2991.* Salem, Oregon: Author, 1995. 26 pages.

Paris, Scott G.; Theresa A. Lawton; Julianne C. Turner; and Jodie L. Roth. "A Developmental Perspective on Standardized Achievement Testing." *Educational Researcher* 20, 5 (June-July 1991): 12-20. EJ 432 070.

Perrone, Vito, editor. *Expanding Student Assessment.* Alexandria, Virginia: Association for Supervision and Curriculum Development, 1991. 168 pages. ED 377 489.

Perrone, Vito. "On Standardized Testing: A Position Paper of the Association for Childhood Education International." *Childhood Education* 67, 3 (Spring

1991): 131-42. EJ 430 344.

Popham, W. James. "Circumventing the High Costs of Authentic Assessment." *Phi Delta Kappan* 74, 6 (February 1993): 470-73. EJ 457 280.

Raths, James; Lilian Katz; and John Fanning. *The Status of Primary School Reform in Kentucky and Its Implications.* In *First Year Reports to the Prichard Committee*. Lexington, Kentucky: The Prichard Committee for Academic Excellence, July 1992.

Rhodes, Lynn K. (editor). *Literacy Assessment: A Handbook of Instruments*. Portsmouth, New Hampshire: Heinemann, 1993. 181 pages.

Seeley, Marcia M. "The Mismatch Between Assessment and Grading." *Educational Leadership* 52, 2 (October 1994): 4-6.

Shepard, Lorrie A. "Using Assessment to Improve Learning." *Educational Leadership* 52, 5 (February 1995): 38-43. EJ 497 535.

_____. "Why We Need Better Assessments." *Educational Leadership* 46, 7 (April 1989): 4-9. EJ 387 134.

Slavin, Robert E. "Learning Together." *American Educator* 10, 2 (Summer 1986): 6-11. EJ 338 410.

Smith, Mary Lee. "Put to the Test: The Effects of External Testing on Teachers." *Educational Researcher* 20, 5 (June-July 1991): 8-11. EJ 432 069.

Stiggins, Richard J. "Assessment Literacy." *Phi Delta Kappan* 72, 7 (March 1991): 534-39. EJ 422 814.

Viadero, Debra. "Even as Popularity Soars, Portfolios Encounter Roadblocks." *Education Week* XIV, 28 (April 5, 1995): 8-9.

_____. "New Assessments Have Little Effect on Content." *Education Week* XIV, 40 (July 12, 1995): 6.

Westmoreland Elementary School. *Menu of Student Characteristics and Questionnaire for Parents of Kindergarten Students and New Students*. Eugene, Oregon: Author, undated a. 4 pages.

_____. "Primary Progress Report." Eugene, Oregon: Author, undated b.

Wiggins, Grant. "A True Test: Toward More Authentic and Equitable Assessment." *Phi Delta Kappan* 70, 9 (May 1989): 703-13.

_____. *Assessing Student Performance: Exploring the Purpose and Limits of Testing*. San Francisco, California: Jossey-Bass Publishers 1993. 316 pages.

_____. "Toward Better Report Cards." *Educational Leadership* 52, 2 (October 1994): 28-37.

Wolfe, Barbara; Michael Dalton; and Wayne Neuberger. *Oregon Statewide Writing Assessment 1991 and 1992, with Student Writings Grades 3 and 5*. Salem, Oregon: Oregon Department of Education, April 1993. 96 pages. ED 366 960.

Worthen, Blaine R. "Critical Issues That Will Determine the Future of Alternative Assessment." *Phi Delta Kappan* 74, 6 (February 1993a): 444-48, 450-54.

_____. "Is Your School Ready for Alternative Assessment?" *Phi Delta Kappan* 74, 6 (February 1993b): 455-56. EJ 457 278.

Worthen, Blaine R., and Vicki Spandel. "Putting the Standardized Test Debate in Perspective." *Educational Leadership* 48, 5 (February 1991): 65-69. EJ 421 353.

Interviews

Merry Denny, Primary Profile Developer, Advanced Systems in Measurement and Evaluation, Inc., 1941 Bishop Lane, Suite 300, Louisville, Kentucky 40218. Telephone interviews June 29 and August 30, 1995

Ric Hovda, Director, Center for the Collaborative Advancement of the Teaching Profession, University of Louisville School of Education, Louisville, Kentucky. Telephone interview January 19, 1996.

Julie McCann, Principal, Fairplay Elementary School, 3855 NE Highway 20, Corvallis, Oregon 97330. Telephone interview August 17, 1995.

Anita McClanahan, Coordinator, Early Childhood Education, Oregon Department of Education, Public Service Building, 255 Capitol Street NE, Salem, Oregon 97310-0203. Telephone interview September 7, 1995.

Mary Nall, Coordinator, Field Services Team, Ministry of Education, 634 Humboldt St., Victoria, BC V8V 2M4. Telephone interviews June 28 and September 12, 1995.

Carol Olson, Primary Teacher, Westmoreland Elementary School, 1717 City View, Eugene, Oregon 97405. Telephone interview August 22, 1995.

Terry Snyder, Primary Teacher, Westmoreland Elementary School, 1717 City View, Eugene, Oregon 97405. Telephone interview August 24, 1995.

Anne Marie Spizzuoco, Teacher, Morse Street School, 21 Morse Street, Freeport, Maine 04032. Telephone interview August 28, 1995.

Cheryl White, Principal, Morse Street School, 21 Morse Street, Freeport, Maine 04032. Telephone interviews August 22 and September 7, 1995.

Barbara Wolfe, Coordinator of Assessment, Oregon Department of Education, Public Service Building, 255 Capitol Street NE, Salem, Oregon 97310-0203. Telephone interview August 1, 1995.

OREGON SCHOOL STUDY COUNCIL

Other Multiage Titles Available from OSSC

• Nongraded Education: Overcoming Obstacles to Implementing the Multiage Classroom

Explores common barriers to implementation of nongraded instructional practices, suggests ways to surmount them, and stresses the need for adequate planning and preparation. (November/December 1994) 84 pages. $11. **OB1294**

• Making the Transition from Graded to Nongraded Primary Education

Focuses on the delicate process of transition from graded to nongraded organization, offering examples at both the local and state level. With its companion Bulletin (March 1992), it provides a comprehensive look at nongraded education at the primary level. (April 1992) 41 pages. $8. **OB0492**

• Nongraded Education: Mixed-Age, Integrated, and Developmentally Appropriate Education for Primary Children

Examines the history of nongraded education, research on its effectiveness, and ways it is put into practice. First of two Bulletins on nongraded primary education. (March 1992) 38 pages. $7. **OB0392**

ORDER

Nongraded Education: Overcoming Obstacles to Implementing the Multiage Classroom	no. of copies ____ x	$11.00	____ . __
Nongraded Education: Mixed-Age, Integrated, and Developmentally Appropriate Education for Primary Children	no. of copies ____ x	$7.00	____ . __
Making the Transition from Graded to Nongraded Primary Education	no. of copies ____ x	$7.00	____ . __
		Shipping $3.00	____ . __
		Total	____ . __

Name: _____

Phone: ( ) _____

Institution: _____

Position: _____

Address: _____

City: _____ *State:* _____ *Zip:* _____

Payment: Full payment or purchase order must accompany all orders. Make checks payable to **University of Oregon/OSSC**.

Quantity Discounts: 10-24 copies of a single title 15%; 25-49 copies 20%; 50+ copies 25%.

Delivery: Shipping is by UPS Ground or equivalent. For faster delivery, call for charges.

Oregon School Study Council, 5207 University of Oregon, Eugene, Oregon 97403-5207. (800) 438-8841 or (541) 346-5044. Fax (541) 346-2334.

1-800-438-8841

Oregon School
Study Council
University of Oregon
5207 University of Oregon
Eugene, OR 97403-5207

Nonprofit
Organization
US Postage
PAID
Eugene OR
Permit No. 63

OSSC BULLETIN

82